Comparative Report

Final Evaluation of the Building Capacity to Use Research Evidence (BCURE) Programme

Date: January 2018

Authors: Isabel Vogel and Melanie Punton

Submitted by Itad

Results in development

Acknowledgements

The authors would like to thank the BCURE evaluation team for their contributions to the synthesis. We would also like to thank the BCURE implementing partners for their invaluable support and engagement during the evaluation process and country visits. The DFID team at Evidence into Action have provided supportive management and helpful guidance, which have considerably strengthened the evaluation.

Evaluation team

The evaluation was undertaken by a core team from Itad, in partnership with Stellenbosch University.

Core team members

Isabel Vogel (Itad Associate)	Team Leader and lead evaluator for Kenya case study
Rob Lloyd (Itad Associate Director)	Project Director
Melanie Punton (Itad Senior Consultant)	Lead evaluator for Zimbabwe case study, support to Bangladesh case study, methods advisor
Gregory Gleed (Itad Consultant)	Lead evaluator for Pakistan case study
Joe Bolger (independent consultant)	Lead evaluator for Bangladesh case study, support to Pakistan case study, lead researcher on impact case study
Teresa Hanley (independent consultant)	Lead evaluator for Sierra Leone case study (Stage 1 and 2), support to Kenya case study
Babette Rabie (Stellenbosch University)	Lead evaluator for South Africa case study (Stage 1 and 2), support to Zimbabwe and impact case studies
Fanie Cloete (Stellenbosch University)	Lead evaluator for India and Zimbabwe case studies (Stage 1), advisory support

Country case studies were supported by national consultants Maheen Sultan (Bangladesh), Alfred Odour and Susan Mathai (Kenya), Rafiq Jaffer (Pakistan), Munhamo Chisvo (Zimbabwe), Andrew Lavalei (Sierra Leone), Benita Williams (South Africa) and Milindo Chakrabarti (India). Research assistance was provided by Alexina Jackson, Greg Smith, Verdiana Biagioni, Louise Horton and Katharine Hagerman. Communications support was provided by Clare Gorman and Emmeline Henderson.

Disclaimer

The views expressed in this report are those of the evaluators. They do not represent those of DFID or of any of the individuals and organisations referred to in the report.

Contents

Click on headings below to navigate

Annexes are at the end of this document, with a separate contents page.

List of acronyms

AFIDEP	African Institute for Development Policy
ACD	Africa Cabinet Decision-Making (ACD) Programme
ACGN	Africa Cabinet Government Network
ASI	Adam Smith International
BCURE	Building Capacity to Use Research Evidence (programme)
CFP	Cabinet Focal Person
CIMO	Context–Intervention–Mechanism—Outcome
CMO	Context–Mechanism–Outcome
(C)ToC	(Common) Theory of Change
DEA	Department of Environmental Affairs
DFID	Department for International Development
DPME	Department for Planning, Monitoring and Evaluation
E&T	Excise and Taxation
EIPM	Evidence-Informed Policymaking
EQ	Evaluation Question
M&E	Monitoring and Evaluation
MoC	Ministry of Commerce
MoEF	Ministry of Environment and Forests
MoH	Ministry of Health
MoHFW	Ministry of Health and Family Welfare
MoYIEE	Ministry of Youth, Economic Empowerment and Indigenisation
NES	National Evaluation System
PEA	Political Economy Analysis
R&D	Research and Development
R4H	Research for Health
SECURE	Strengthening Capacity to Use Research Evidence
UJ	University of Johannesburg
UK	United Kingdom

Executive Summary

> “Evidence is crucial to successful policymaking. However, in many low and middle-income countries, policy makers lack the capacity to effectively access, appraise and apply research when making decisions.”

This was the starting assumption behind the Building Capacity to Use Research Evidence (BCURE) programme – a £15.7 million initiative funded by the UK Department for International Development (DFID) from 2013–17. This report presents the findings of the three-year realist evaluation of BCURE.

Headline Findings

Working with governments to build capacity for evidence use requires a politically informed and multidimensional approach. Capacity gaps should be viewed as just one element of a tapestry of factors that block or disincentivise evidence-informed policymaking.

First, evidence use is inherently political. It is often constrained in low and middle-income countries by authoritarian, politicised and fragmented institutions, which are hobbled by financial constraints, low technical or policy experience among civil servants and high levels of corruption. Despite these challenges, many countries are embarking on reforms that create momentum for evidence-informed policy. Building capacity for evidence use means thinking and working politically to harness these windows of opportunity, and effectively navigating political economy constraints that can undermine meaningful reform. Second, changing ways of working requires thinking beyond ‘skills’ to build capacity at multiple levels of complex government systems. Individual capacity (in terms of knowledge, skills, confidence and commitment) is the bedrock of effective evidence use, but programmes also need to harness organisational processes, management support and wider incentives for people to change ways of working, and make sure interventions join up to have a system-wide effect. Finally, external partners should accompany change, not impose it. Government reform processes are unpredictable and highly context-specific, meaning that it is rarely clear at the outset what will work. Success is more likely when programmes accompany government partners through a process of change in a flexible, tailored and collaborative way, rather than providing ad hoc support through one-off activities.

About BCURE

BCURE consisted of six linked capacity building projects across 12 low and middle-income countries in Africa and Asia. Each used different interventions to build capacity for evidence use, designed and combined in different ways – including training, mentoring, technical support to develop evidence tools and guidelines, learning exchanges and policy dialogues. Projects ranged in scope and scale, from working in single ministries to working across whole government systems. The evaluation focused on BCURE's work in six countries: Bangladesh, Kenya, Pakistan, Zimbabwe, Sierra Leone and South Africa.

About the evaluation

The evaluation used a realist approach to explore how and why capacity building for evidence-informed policymaking works and does not work, for whom, to what extent, in what respects and in what circumstances. It encompassed annual internal programme evaluations of the six BCURE projects, a literature review, an impact case study of a non-BCURE capacity building initiative, and annual synthesis reports. This final report summarises insights from across these components. Findings are based largely on qualitative interviews with more than 500 stakeholders over three years, including BCURE programme staff, participants in BCURE activities, non-participating colleagues and managers, high-level government officials, and civil society stakeholders. The evaluation also draws on BCURE monitoring data and programme documentation, and where possible government documents such as policy products.

The evaluation used a realist approach to explore how and why capacity building for evidence-informed policymaking works and does not work, for whom, to what extent, in what respects and in what circumstances

Evaluation key facts

3 years

3 stages of data collection and analysis

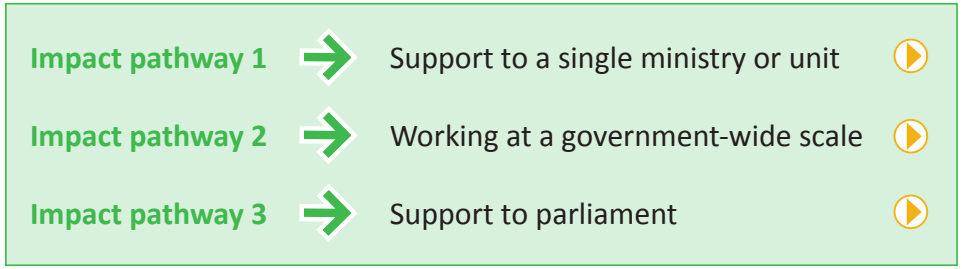
6 countries

15 country reports

567 stakeholders consulted

What did BCURE achieve?

BCURE took three main routes to promoting evidence-informed policy making, at different levels of government. We have termed these 'impact pathways', which tell the story of how and why BCURE projects influenced a sequence of outcomes towards the desired impact. We hope these can provide a broad, non-prescriptive road map for future programmes working to build capacity for evidence use in government settings.



It is too early to judge whether any impact pathway is 'better' than others. BCURE was only four years long – a very short time to observe change in government behaviour and processes, and too early to assess how far BCURE contributed to a step change in the use of evidence. The key lessons below suggest that the choice of approach should be based on an assessment of where political dynamics offer potential to catalyse change, and where existing relationships and networks can give an external partner a 'way in.'

The evaluation noted three different levels of success across the impact pathways:

- **Significant progress towards catalysing change at scale: Bangladesh and Sierra Leone (cross-government) and Kenya (Parliament).** These projects involved 'top down' activities to establish procedures and incentives for evidence use at an organisational level, combined with 'bottom up' capacity building for technical staff – a model that appears to have significant potential to catalyse long-term progress towards improved evidence use. However, the findings are tentative across all three settings, and although there is strong evidence of senior-level ownership it is too early to tell whether early progress will continue. This depends on continued political leadership, high-level incentives and resource mobilisation – all potentially fragile and subject to change.
- **Pockets of success around specific policy processes and capacitated units in single ministry settings: Kenya, Zimbabwe, Pakistan and South Africa.** Across most BCURE projects, there are examples of improved capacity at an organisational level, or good quality policy pilots or tools that have showcased the value of evidence use and have been adopted by a government unit. This happened where projects identified clear windows of opportunity and provided collaborative support within settings where there were existing incentives for change. However, these examples did not add up to system-level change to embed evidence use, which may impede their long-term influence.
- **Ad hoc and 'one dimensional' change: all six projects and all three impact pathways.** Across the BCURE portfolio, there are many examples of individuals applying new knowledge and skills within specific policy processes as a result of capacity support from BCURE – but while these are important demonstrations of individual behaviour change, they are ad hoc and unlikely to add up to a step change. At an institutional level, in Pakistan, training on evidence-informed policymaking was adopted into national civil service training – but while this was a significant achievement, it was not joined up to other activities or broader reforms so has limited chance of contributing to a step change on its own.

BCURE was only four years long - a very short time to observe change in government behaviour and processes, and too early to assess how far BCURE contributed to a step change in the use of evidence

How and why can capacity building improve evidence-informed policy making?

Programmes aiming to build capacity for evidence use are often designed around specific activities, such as training or technical support. Our findings suggest the need to instead begin by considering the key ways of working and mechanisms (or change processes) that underpin success – and then think through how best to catalyse these in a particular context. We have identified six lessons on how and why capacity building can improve evidence-informed policymaking:

Lesson 1: BCURE highlights the importance of thinking and working politically

All six BCURE projects were superficially a good fit with government agendas around evidence-informed policy making, with some level of demand from senior leaders, and were tailored to align with ministry requirements through needs assessments. However, scoping activities should have looked beyond 'face value' statements of interest, and considered deeper internal political economy dynamics within ministries which shaped the potential for catalysing change.

BCURE had greater success in catalysing the key mechanisms where partners located an entry point in a sector or government institution where there was existing interest in evidence, clear incentives for reform, and a mandate for promoting evidence use; took advantage of a window of opportunity for partnership and reform, often building on existing institutional credibility and relationships to gain a foot in the door; and nurtured relationships with individual champions who acted as 'gatekeepers' and 'cheerleaders' for the programme.

Mechanisms underpinning success in BCURE

Accompaniment: where an external partner provides tailored, flexible and responsive support to a government institution through a process of reform, characterised by a high level of trust.

Self-efficacy: where providing information, opportunities to practise skills, coaching or technical support builds individuals' confidence in their ability to do their jobs or achieve a particular goal.

Facilitation: where an evidence tool, system or process facilitates government officials to do their jobs or undertake a task more easily or efficiently.

Reinforcement: where rewards or other forms of control create incentives that motivate officials to work in a particular way. Positive reinforcement includes rewards and encouragement, while negative reinforcement includes reminders, audits and mandatory requirements.

Showcasing: where good examples of evidence tools or processes demonstrate the value of an evidence-informed approach, which leads to them being adopted elsewhere.

Adoption: where senior government stakeholders decide to adopt a new evidence tool, system or process to help standardise evidence-informed policymaking within a government institution. This can be on a small or a large scale.

Critical mass: where changes in practice among a sufficient number of government officials diffuse out to influence colleagues' behaviour, and the rate of adoption of new behaviours becomes self-sustaining.

Lesson 2: Programmes should accompany change, not impose it

BCURE had most success where projects 'accompanied' government partners in a flexible, tailored, collaborative way that promoted ownership, and strengthened partner capacity through 'learning-by-doing'. This was possible where government partners already had a mandate to promote evidence use, or where BCURE had built up relationships and trust through previous activities that led to an invitation to accompany policy processes.

Accompaniment is not straightforward, and projects are likely to face numerous blockages that need to be navigated, including frequent staff rotations, corruption scandals, and changes in government priorities. In order for programmes to work in this way, there needs to be sufficient flexibility in the contracting model, to allow partners to respond nimbly to challenges and opportunities.

BCURE had most success where projects 'accompanied' government partners in a flexible, tailored, collaborative way

Lesson 3: Changing behaviour requires more than building skills through training

All six BCURE projects used training as a key intervention, but there were marked differences in the extent to which trainees were able to apply their learning. Where BCURE led to more routine changes in evidence access, appraisal and use, this was often because projects succeeded in catalysing multiple mechanisms together: building self-efficacy, providing tools that facilitated staff to do their jobs more easily, and tapping into or generating organisational incentives to reinforce behaviour change. In many cases, training did not lead to change in practice as a result of a broader environment unconducive to evidence-informed ways of working, and issues with training design and implementation.

The evaluation highlights the importance of following good practice in adult learning theory. Behaviour change is more likely where activities are closely targeted to individuals who can apply their learning because it is directly relevant to their day-to-day work, where follow-up support helps embed learning, and where training is practical and participatory, uses local case studies or live policy examples, and is delivered by good quality facilitators who understand the specific sector as well as the broader national context.

Lesson 4: Catalysing a 'critical mass' of evidence users requires specific and targeted strategies

A common assumption in BCURE was that training a 'critical mass' of individuals would diffuse out to influence broader change. However, even where BCURE succeeded in changing behaviour, there is limited suggestion that this influenced people's colleagues or managers. The evidence relating to this mechanism is therefore limited, although some tentative lessons can be inferred.

First, if training is not directly relevant or there are missing incentives and organisational structures to support evidence use, then individuals may be unable to apply their learning in the first place, and so there is little prospect of them influencing others. Second, if individuals are too scattered across siloed units and divisions then this dilutes their opportunity to influence. Third, if officials are based in a unit that has limited power and resources, or if the programme works only with technical staff but not their managers or senior decision makers, this limits the possibility of influencing senior-level attitudes or behaviours.

Building in an explicit 'training of trainers' strategy, supported by a 'clustering' approach where individuals from the same unit are targeted, may help trainees develop social connections to provide mutual support, or act as a 'focal point' for promoting new ways of working.

Lesson 5: Supporting practical tools or policy pilots can showcase the value of evidence

Several BCURE partners provided practical support to policy processes, or helped develop tools that enabled officials to engage with evidence more easily. This proved one of the most successful interventions, leading to new tools and evidence-informed policies in Kenya, Bangladesh, Pakistan and South Africa. Success was due to BCURE partners identifying an entry point where there was a real need to solve a policy or service delivery problem, and the potential to build on existing work and partnerships and leverage external resources. It proved essential to secure high-level support for the process, and involve stakeholders at the right level of seniority and with the right technical and interpersonal skills, from within and outside government.

Where BCURE provided hands-on support to co-produce policies and tools (through a model of 'accompaniment'), rather than developing the tools themselves, this helped ensure ownership and in turn made adoption of the resulting tool or process more likely – and it also supported 'learning-by-doing,' helping participants embed skills gained through training.

Supporting practical tools or policy pilots proved one of the most successful interventions, leading to new tools and evidence-informed policies in Kenya, Bangaldesh, Pakistan and South Africa

Lesson 6: Promoting 'genuine' adoption of reforms is essential for sustainable change

Where a programme aims to build capacity for evidence use, the goal should be to promote formal adoption of a new process, tool or practice, and ensure it is supported and resourced by senior managers – in order for this to continue incentivising behaviour change once the programme ends.

BCURE succeeded in catalysing 'small-scale adoption' of new tools or guidelines in specific units or sectors, including in South Africa and Pakistan. In these countries tools proved genuinely useful to officials' work, senior managers could see their value, and there was a clear institutional home for the tools as well as resources for scale-up.

'Large-scale adoption' involved rolling out a new system or process on a government-wide scale. For example, courses on effective evidence use were adopted into training institutes in Bangladesh and Pakistan, through the support of high-level champions in contexts with an established culture of civil service training. However, embedding training in this way carries risks, as it dilutes the factors found to catalyse individual behaviour change through shortening training courses, watering down the targeting, and stripping out follow-up support.

BCURE also catalysed large-scale adoption of government-wide tools and procedures to support evidence use in Bangladesh and Sierra Leone. However, there is a real risk in both countries that adoption will happen on paper but not in practice. 'Genuine' adoption requires ongoing government ownership and access to resources beyond the project, as high-level incentives shift and new opportunities rise and fall in dynamic political environments.

Building Capacity to Use Research Evidence (BCURE): what worked? Lessons from the evaluation

Link

1

"Evidence is crucial to successful policy making. However, in many low and middle-income countries, policy makers lack the capacity to effectively access, appraise and apply research when making decisions."

This was the starting assumption behind BCURE. The evaluation investigated how, why and for whom capacity building worked and didn't work across the BCURE projects.

2

In order to build capacity for evidence use we found **three 'ways of working'** that underpined success in BCURE:

Thinking and working politically

Programmes need to understand the political and power dynamics that affect evidence use in government.

Accompanying rather than imposing change

BCURE was more successful when partners 'accompanied' government partners through a flexible, tailored, collaborative approach that promoted ownership.

Working at multiple levels of the system

Individual capacity is the bedrock for effective evidence-informed policy making – but programmes also need to strengthen systems, develop tools, and nurture champions.

3

The BCURE partners took **three broad entry points** - or impact pathways - to working with government on evidence-informed policy making.

- Single ministry or unit
- Across government
- Parliament

4

Across the three impact pathways, success followed when BCURE managed to activate a combination of 'mechanisms' (change processes). These led to changes in skills, attitudes, behaviour and systems, which laid the foundations for more routine use of evidence in government.

Self-efficacy

Example: In Zimbabwe, training built officials' confidence to use evidence in the Ministry of Youth, helping them work more effectively in their new roles as officers in a recently-established research unit.

Critical Mass

Example: In Kenya, officials in the Ministry of Health cascaded their learning from BCURE through adapting the training curriculum, mobilising financial support from a separate funder, and training county level policy makers in evidence-informed policy making.

Reinforcement

Example: In Sierra Leone, BCURE supported new Cabinet-level processes and templates, making it mandatory for line ministries to consider evidence in policy submissions. A new unit with the mandate to follow up on implementation created further pressure to comply.

Outer circle represents outcomes that are important to drive routine evidence use, but were not a core focus of the evaluation

Showcasing

Example: In South Africa, BCURE helped produce an 'evidence map' that gathered together diverse sources relating to human settlements. Learning was shared through reports and workshops, leading to demand for further maps by various ministries.

Adoption

Example: In Bangladesh, BCURE piloted an evidence training course that was adopted nationally and will reach thousands of civil servants each year. BCURE also co-developed evidence-informed policy making guidelines, which have been adopted by Cabinet with the intention of rolling them out across all government ministries.

Facilitation

Example: In Pakistan, BCURE developed data visualisation tools to help front line service providers understand what was happening on the ground – for example a dashboard showing tax collection by area, which helped officials manage staffing and performance.

1. Introduction

This report presents summative findings from the independent realist evaluation of the Building Capacity to Use Research Evidence (BCURE) programme. It explores how and why capacity building for evidence-informed policymaking (EIPM) works and does not work, for whom, to what extent, in what respects and in what circumstances.

The £15.7 million BCURE programme aimed to improve the use of evidence in decision making in low and middle-income countries. It ran from 2013 to 2017, funded by the UK Department for International Development (DFID), and was made up of six linked projects implemented across 12 countries in Africa and Asia, as well as a number of additional countries reached through international networking and small-grant initiatives. Each project used different combinations of capacity development interventions to support policymakers, government officials and parliamentarians to develop skills, knowledge and systems in order to improve the use of evidence in decision making.

The BCURE evaluation was funded by DFID, conducted by an independent evaluation team from Itad, and ran from 2014 to 2017, in parallel with the programme. The evaluation had a focus on both learning and accountability, as expressed in its two key aims (see Annex 1):

- To strengthen the global evidence base on the effectiveness of capacity building approaches to support evidence-informed policy.
- To evaluate the effectiveness and value for money of the six BCURE programmes.

The evaluation encompassed annual internal programme evaluations of the six BCURE programmes, a literature review, an impact case study of a non-BCURE capacity building initiative, and annual synthesis reports on how and why capacity building for evidence use works or not in different contexts. This final report summarises insights from across all of these components.

The main intended users are BCURE's managing team at DFID's Research and Evidence Division and the partners responsible for delivering BCURE, to inform future programmes. Insights are also of relevance to various other DFID-funded programmes, including national and sector-specific public sector reform and service delivery support interventions, programmes developing evidence bases of 'what works' in various sectors, and 'think tank' and research systems strengthening programmes. Finally, we hope the findings will be of interest to a wide audience of donors, funders, commissioners and implementers working on promoting capacity for EIPM in numerous fields, including research and evidence utilisation, governance, and public management and administration. Further details on key audiences and the communications strategy are contained in Annex 3.

The report does not provide a performance assessment of the BCURE projects individually, or of BCURE as a whole. It draws primarily from case studies of BCURE's work in six of the twelve focal countries, selected to reflect a range of more and less favourable settings for EIPM. This allowed in-depth analysis of how and why the programme has and has not worked in specific contexts, but does not reflect the whole of what has been achieved across the programme. It also does not provide a holistic story of each of the BCURE projects. Project performance has been discussed in annual programme evaluations which are internal to DFID – headline findings are summarised in published BCURE Annual Reports available on the DFID website.

This report is not a 'typical' evaluation report, with findings structured against evaluation questions. It is a synthesis product, drawing on a large number of detailed country-level reports conducted over three years. In order to draw out the key messages from this complex programme as clearly as possible for the readers, we have identified three 'impact pathways' that represent

specific routes taken by BCURE projects to work within government: single ministry, cross-government, and support to Parliament. The impact pathways present the sequence of outcomes, mechanisms, key approaches and interventions that underpinned success in the BCURE programmes. The report is structured around these impact pathways, providing a rich and context-specific explanation of how and why capacity support through BCURE has (and has not) promoted evidence use through entry points at different levels.

The report is structured as follows:

- **Section 2** presents the evaluation approach and methodology.
- **Section 3** describes the six BCURE projects, the country contexts in the settings visited for the evaluation, and the evaluation's theory about how and why BCURE was expected to lead to change, including an explanation of the three impact pathways.
- **Section 4** outlines three cross-cutting ways of working that were found to underpin success across BCURE, drawing links to relevant wider literature.
- **Section 5**, **Section 6** and **Section 7** present the evaluation findings, structured in line with the three impact pathways. These sections represent the core of the report, where the evidence on how far, how and why BCURE projects contributed to improved EIPM is unpacked and explored in depth.
- **Section 8** concludes, and presents the evaluation's revised and tested theories about how and why capacity building can contribute to EIPM.
- **Section 9** draws out lessons for future programmes.

Throughout the report, various 'spotlight' sections also draw out cross-cutting findings from the BCURE portfolio.

- **Spotlight 1.** BCURE's approach to gender, equity and social inclusion
- **Spotlight 2.** How and why can training support evidence informed policymaking?
- **Spotlight 3.** Creating spaces for conversation through networks, policy dialogues, knowledge cafes and learning events
- **Spotlight 4.** Establishing a sustainable national actor through 'learning-by-doing'

2. Evaluation design and methodology

The evaluation design and methodology is summarised below. Further details are provided in Annex 3.

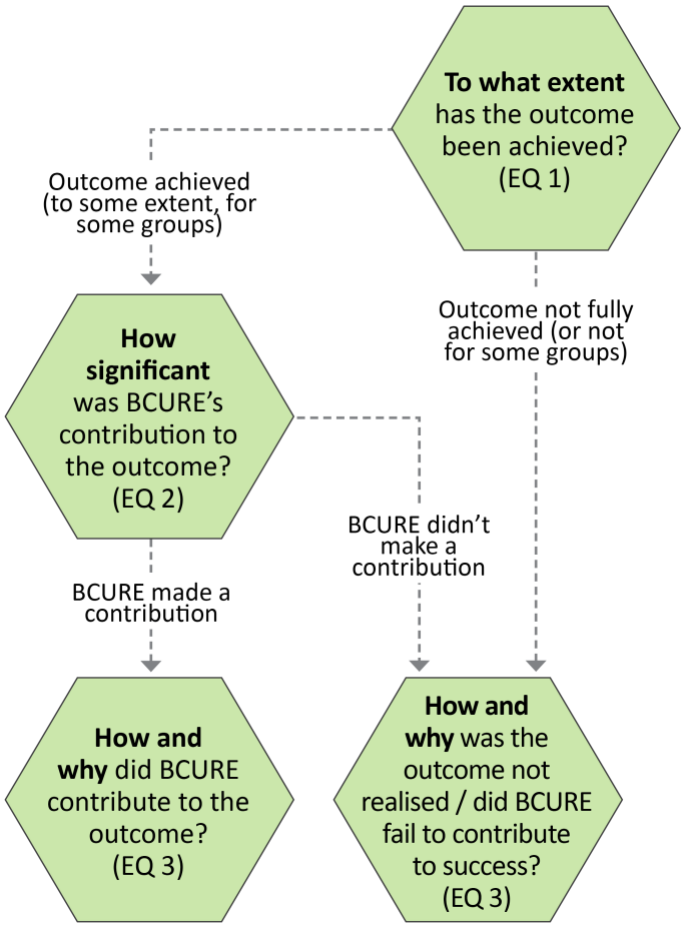
2.1 Evaluation questions

The BCURE evaluation addresses two overarching evaluation questions (EQs). These are based on the questions posed in the Terms of Reference (Annex 1), revised in the inception phase following discussions with DFID.

1. How effective are the BCURE projects in achieving their stated outcome of increasing the use of evidence in public sector decision making, and influencing longer-term changes in policy quality?
2. How and why does capacity building for evidence use work and not work, for whom, to what extent, in what respects and in what circumstances?

More specific evaluation questions have been developed at each stage of the evaluation (see Annex 3), adjusted in line with revisions to the design. At Stage 3, it was decided to focus on a smaller number of 'priority outcomes' rather than investigate all of the anticipated outcomes across the BCURE projects. The revised EQs were as follows:

Figure 1. Logical flow of the EQs



To what extent have priority outcomes been realised and for whom, and how sustainable are they?

1. Have the theorised changes happened?
   - How far have these changes occurred across different sub-groups and organisations etc., reflecting on gender and equity issues?
   - How sustainable are the changes?
2. How significant was BCURE's contribution to priority outcomes, alongside the contribution of non-BCURE factors?
   - What is the evidence that BCURE contributed to causing the observed changes, and what is the evidence that non-BCURE factors contributed?
   - What is the relative importance of BCURE and non-BCURE factors in explaining the observed changes?

3. How and why did BCURE contribute or fail to contribute to priority outcomes?
    - Through which mechanisms, enabled by which features of the intervention and features of the (individual, interpersonal, organisational and institutional) context, did BCURE contribute to the observed changes?

To answer the three EQs, the Stage 3 evaluation gathered and analysed evidence from various sources against country-level theories of change, to judge the extent to which an expected outcome had emerged (EQ 1), the extent to which BCURE contributed to this outcome (EQ 2), and how, why, for whom and in what circumstances the outcome had and had not happened (EQ 3). Figure 1 depicts the logical flow of the evaluation questions. As agreed with the evaluation Steering Committee, the evaluation questions were framed around case-specific priority outcomes and thus were answered at the level of the internal country case study reports. This overview report provides summary comparative judgements across the cases in relation to the EQs, but its purpose is not to answer the questions at a portfolio level.

2.2 Approach to answering the evaluation questions

BCURE worked in complex government contexts, with myriad contextual conditions influencing potential outcomes. These included diverse historical institutional trajectories; varied political and economic conditions, government systems and organisational cultures; and a wide range of participant characteristics (individuals' identities, gender and ethnicities). Quasi-experimental and counterfactual approaches are unsuited to evaluating this type of programme, as there is no possibility of establishing a control group or comparator (Stern et al., 2012). In addition, BCURE was likely to be just one of a number of factors influencing change in complex government systems, giving rise to the 'attribution problem' – the challenge of attributing a particular change to a particular programme when other factors are also contributing (Wimbush et al., 2012).

In order to address these challenges and answer the evaluation questions, the evaluation adopted a **realist evaluation** approach, drawing on elements of **contribution analysis** and taking a **political economy lens.**

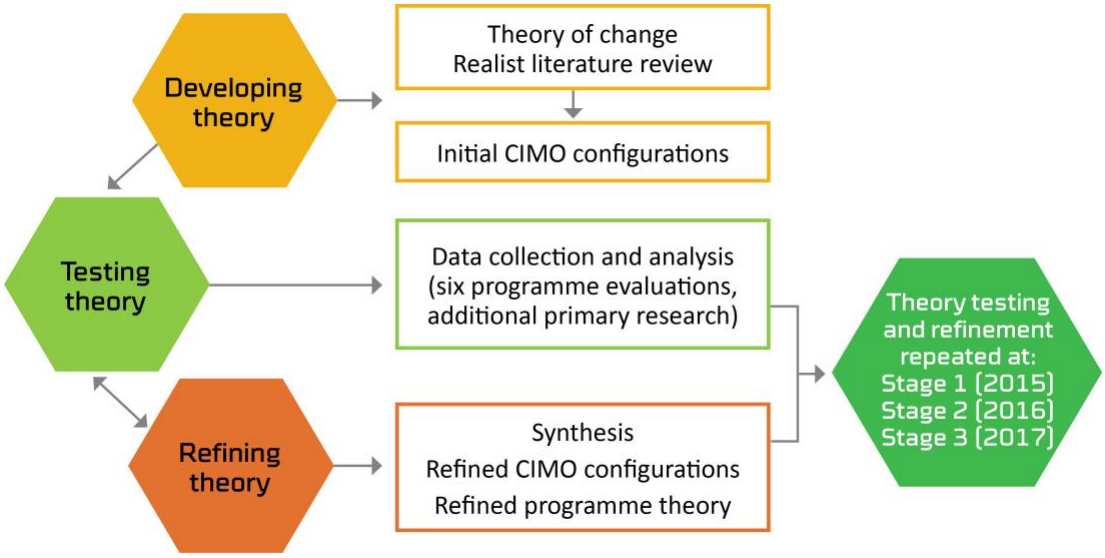
Realist evaluation

A realist approach was selected because the primary aim of the evaluation was to strengthen the evidence base on how capacity building can promote EIPM, to inform decisions within and beyond DFID about whether to fund and how to design this type of programme in future. DFID was interested in understanding not just *whether* BCURE worked but also *how and why* capacity building can contribute to increased use of evidence in policymaking in the very different contexts in which the programme is operating (EQ 3). Realist evaluation meets these objectives through opening up the 'black box' between BCURE interventions and outcomes, by developing and testing *programme theory* (an explanation of how, why and in what contexts interventions lead to particular outcomes – see Box 1).

Programme theory consists of linked sets of hypotheses about the *mechanisms* that cause an intervention to work or not work in particular *contexts*, to lead to specific *outcomes*. These hypotheses are known as 'context–mechanism–outcome' or CMO configurations (see Box 1) – the core analytical units of realist evaluation (Pawson and Tilley, 1997; Wong et al., 2013). The evaluation team decided to incorporate *features of the intervention* as an additional element to our CMO configurations, in order to separate out features that are inherent in or under the control of the programme (such as training design or length) from contextual factors that are not (such as professional incentives to participate in training) when considering what might 'spark' a particular mechanism. This gives us the formulation C+I+M=O (CIMOs), used throughout this report.

Realist evaluation encompasses three broad stages: developing theory, testing theory and refining theory. These are iterative rather than linear; theory is developed, tested, refined and tested again as knowledge accumulates. Figure 2 provides an overview of the evaluation design.

Figure 2. Application of a realist approach in the BCURE evaluation



Box 1: Context, mechanism, outcome and programme theory

Mechanisms are the causal forces, powers, processes or interactions that generate change within an intervention – including the choices, reasoning and decisions people make as a result of the resources the programme provides. An intervention such as a training course is not a mechanism. The mechanism is the 'thing' that explains *why* training changes behaviour (or does not) in a particular setting.

Mechanisms are triggered only in certain **contexts**. Contextual factors may include *individual* characteristics that affect how people respond to opportunities (e.g. gender, ethnicity, education); *interpersonal* factors that affect trust and buy-in (relationships between stakeholders and programme implementers); *institutional* factors (the rules, norms and culture of the organisation in which the intervention is implemented); and *infrastructural* factors – the wider social, economic, political and cultural setting of the programme (Pawson and Tilley, 2004).

Outcomes refer to intended and unintended short-, medium- and long-term changes resulting from an intervention.

A **CMO configuration** is a theory or hypothesis about how a particular mechanism works in a specific context to lead to an outcome. They can usually be read as sentences – for example, *'Where training content is directly relevant to a person's day job (C), providing information about how evidence can improve policymaking can spark an "eye-opener" in which trainees recognise how evidence can add value (M), leading to increased use of evidence in their day-to-day work (O)'.*

A **realist programme theory** explains *'(some of) how and why, in the "real world", a programme "works", for whom, to what extent and in which contexts'* (Wong et al., 2016). A realist programme theory is a variation on a theory of change (ToC), which explicitly spells out the causal links between outcomes as CMO configurations. Factors that would be presented as 'assumptions' in a traditional ToC are embedded into CMOs, as contextual factors and/or conditions necessary for mechanisms to fire. Some ToC approaches also include 'risks to assumptions' – that is, factors that will prevent the assumptions from holding true. Again, realist programme theory integrates this into the CMO testing, by explaining the contextual or intervention factors that block mechanisms from operating.

Sources: Pawson and Tilley, 1997; Westhorp, 2014; Punton et al., 2016b

The first iteration of the BCURE theory drew on the evaluation team's existing knowledge and professional hunches about the nature of capacity building, and how capacity building can contribute to evidence use in policymaking. This was used to shape the research questions for the BCURE literature review, which identified additional theories in the wider literature about how capacity building can contribute to EIPM. These were used to develop our first iteration of CIMO configurations. Stages 1 and 2 of the evaluation then began to test and refine these CIMOs, contributing to a revised programme theory at each stage. At Stage 3, a prioritised set of theories have been tested and revised for a final time, and are presented in this report. Annex 4 contains a full explanation of how the BCURE theory has evolved over time, and lists the CIMOs tested at Stage 3.

Contribution analysis

In order to answer EQ 2, the Stage 3 evaluation drew on elements of contribution analysis – a theory-based evaluation approach that provides a systematic way to arrive at credible causal claims about a programme's contribution to change. This allows a robust assessment of cause and effect when it is not practical to design an experiment to measure the attribution of a particular change to a particular programme (Mayne, 2012). The six steps of contribution analysis[1] provided a framework to help prioritise outcomes and causal links to investigate during Stage 3, and assess the contribution of the programme alongside the role of other factors, as follows:

- A country-level ToC was developed for each case study, allowing the underlying causal logic to be unpacked.
- Evidence from earlier stages of the evaluation was assembled, in order to assess the strength of the existing contribution story, and identify weaknesses and gaps.
- Priority outcomes and causal links to focus on at Stage 3 were then selected, based on a consideration of their importance to the overall contribution story, and utility and importance to stakeholders (Lemire et al, 2012).
- Evidence about the extent of BCURE contribution was then collected through country case studies, including through incorporating questions about contribution in the interview topic guides, and examining other explanations for observed outcomes through the political economy lens.
- The country case study analysis then involved a systematic assessment on the extent of BCURE contribution against the country-level ToCs, described further below.

Political economy lens

The Stage 3 evaluation aimed to incorporate a stronger understanding of how political economy issues affect evidence use in policymaking, in order to unpack non-BCURE drivers of outcomes (EQ 2) and incorporate political economy dimensions into our explanations of why BCURE contributed or failed to contribute to outcomes – that is, the 'C' in CIMOs (EQ 3). A light touch political economy analysis (PEA) exercise was conducted at both country-level (to identify key overarching factors and trends influencing policymaking and evidence use) and sector (looking at the sectors targeted by the BCURE partner), as part of each country case study. This was guided by a framework incorporating a checklist of PEA questions, drawing from various pragmatic PEA tools. The questions were used to structure an initial review conducted by the national consultant prior to data collection, drawing on secondary data sources (see Annex 5). Further information was then collected through primary interviews with sectoral experts and government stakeholders during the main data collection stage.

[1] These six steps are: setting out the cause-effect issue to be addressed; developing a theory of change; gathering existing evidence on the theory of change; assembling and assessing the contribution story and challenges to it; seeking out additional evidence; revising and strengthening the contribution story (Mayne, 2011).

2.3 Evaluation components

The evaluation had four main components:

1. **Annual programme evaluations of BCURE-funded projects**, incorporating primary data collection within one country (the 'country case study'), and analysis of monitoring and implementation documents from all country contexts. At Stage 3, the evaluation refocused its resources to conduct four evaluations instead of six, and limited its analysis to these four countries only. This allowed the team to investigate a smaller number of priority outcomes in more depth.
2. **A realist literature review**, synthesising published papers and grey literature related to capacity building for EIPM.
3. **An impact case study**, consisting of additional primary research on a similar intervention to BCURE that had been running for a longer period of time, in order to provide evidence on how capacity building for EIPM contributes to longer-term improvements in policy quality (the ultimate goal of the BCURE programme).
4. **A synthesis of findings**, drawing together insights on how and why capacity building for evidence use works or does not work in different contexts.

Data collection and synthesis was repeated each year for three years to enable the evaluation to track programme results over time, and iteratively test and refine our theories about how and why particular outcomes have occurred in different contexts – see Figure 1 above. The four components are described in more detail below.

Component 1. Programme evaluations and country case studies

During Stage 1 and 2 of the evaluation, programme evaluation reports were produced for each of the six BCURE projects.[2] At Stage 3 it was agreed with the Steering Committee that the evaluation would conduct four 'country case studies' instead of six programme evaluations, to enable a focus on 'depth' rather than 'breadth'. The reports performed two functions:

1. **Providing internal management reports for each project**, which verified outcomes identified by the BCURE programme monitoring data (and identified additional outcomes); generated an assessment on programme effectiveness, value for money, sustainability and programme contribution to change; and captured key lessons and recommendations in order to inform decision making.
2. **Collecting data on how and why BCURE projects contributed to different patterns of outcomes**. This data was then fed into the synthesis, in order to identify, test and refine theories about how and why capacity building does and does not lead to change.

Each programme evaluation/country case study consisted of an independent review of secondary monitoring data and implementation documents produced by the project team, and primary data collection by the evaluation team within one of the countries targeted by the project. Over the course of the evaluation, 15 programme reports have been produced (five programme evaluations at Stage 1, six at Stage 2, and four country case studies at Stage 3). These are all internal to DFID.

[2] Only five reports were produced at Stage 1, as the Ecorys project had not yet commenced.

Country selection

Case study countries were selected using case replication logic (Yin, 2003), based on a high-level mapping looking at how far BCURE target countries were favourable or challenging environments for EIPM. The selected country case studies represented a range of more and less favourable contexts (detailed in Annex 3.4.2). Pragmatic considerations of security and access also informed the final selection. At Stage 3, the evaluation focused on four of the original six countries, in order to allow a more in-depth exploration.

Country-level theory of change

At Stage 3, country-level theories of change (ToCs) were used as an overarching frame for each country case study. The ToCs provided a focal point for identifying priority outcomes and sectors for investigation at Stage 3, helped identify political economy factors that may affect outcomes, and then provided a structure to the data collection and analysis. This focused on exploring the extent to which priority outcomes in the ToC had been realised or not (EQ 1), BCURE's contribution to these outcomes (EQ 2), and testing our theories (CIMOs) about how and why BCURE contributed or failed to contribute (EQ 3).

Sampling approach

In a realist evaluation, decisions about sampling are driven by a consideration of who the researchers need to talk to in order to test their theory. Our sampling approach was therefore purposive, and iteratively developed between Stages 1 and 3 and within stages based on the evolving programme theory and CIMOs. Respondents were identified using previous samples, stakeholder lists, monitoring data, staff recommendations, and snowball sampling (through recommendations of interview respondents), according to their relationship to BCURE, their role in the government system, their ability to comment on our CIMOs and their relationships to each other. We aimed to triangulate evidence across a range of different stakeholders, through comparing insights from project participants with insights from knowledgeable 'outsiders', and through interviewing participants along with their line managers or colleagues. Each programme evaluation consulted up to 30 stakeholders at each of Stages 1 and 2, and each country case study consulted up to 60 stakeholders at Stage 3. In total, 567 stakeholders were consulted across six countries and over three years. Respondents included programme staff and implementing partners, participants in BCURE activities and non-participating colleagues and managers, high-level stakeholders with an insight into how the government system operates, and stakeholders from civil society and other external vantage points. Many stakeholders were consulted over multiple years, allowing change to be tracked over time. See Annex 3.4.3 for more details.

Data collection sources and methods

The country case studies drew on exploratory workshops with BCURE implementing partner staff, semi-structured interviews with key stakeholders, and independent reviews of programme monitoring data and implementation documents produced by the BCURE partners. This documentation included pre- and post-training course test data, participant feedback on various programme activities, Memoranda of Understanding with government partners, activity reports, meeting minutes, and case studies written by BCURE partners.

Where possible, the evaluation also drew on relevant government documentation, such as policy documents, tools and policies developed in collaboration with BCURE partners, and examples of evidence application by programme participants.

Data analysis methods

Primary and secondary data was extracted into various analysis spreadsheets in Microsoft Excel, which facilitated the coding of data by EQ and CIMO. This provided a systematic catalogue of evidence, enabling country case study leads to transparently assess the strength of evidence behind particular changes, make a judgement about BCURE contribution, and identify how and why changes were thought to have come about through developing and testing CIMOs. The analysis involved systematic comparison of the views of different stakeholders, including programme participants and non-participants, senior and junior staff, and programme managers and civil society. To aid the analysis and to ensure consistency in judgements across the programme evaluations, the programme evaluation leads applied rubrics to assess the strength of evidence underpinning assessments of change and a qualitative judgement on the programme's contribution to change (described further below and in Annex 3.8).

Value for money

A value for money (VfM) analysis was conducted within each country case study, focusing on cost-effectiveness – the extent to which the investments made had delivered value. However, due to the nature of BCURE financial reporting, in many cases it was not possible to identify the precise costs of programme activities. As a result, the cost data was an estimation developed in consultation with BCURE partner staff. Qualitative judgements were made through considering two questions. Firstly, did the outcomes that were achieved justify the costs? Secondly, was the balance of investments across the priority outcomes appropriate?

Component 2. Literature review

A realist literature review (Punton et al., 2016a) was conducted during the early stages of the evaluation, in 2014–15.[3] The findings informed the development of the first iteration of CIMOs tested in Stage 1. A light touch literature review refresh was conducted in 2017 in order to generate additional insights on the Stage 2 programme theory, and the insights have been incorporated into this report.

[3] Available from http://www.itad.com/knowledge-products/bcure-literature-review/

Component 3. Impact case study

The impact case study aimed to generate evidence on how capacity building for EIPM can lead to improvements in the quality of policy processes, the hoped-for ultimate impact of the BCURE programmes. This was designed to complement the BCURE programme evaluations through examining a non-BCURE capacity building intervention: the South African National Evaluation System (NES). Because this had been operating for a longer period of time, it offered the potential to investigate how capacity building (provided by the custodian of the NES: the Department for Planning, Monitoring and Evaluation or DPME) had contributed to changes in policy quality in the longer term. The case study followed the same data collection and analysis methods as the Stage 2 programme evaluations, and took place during Stage 1 and 2 of the evaluation. Some 39 interviews were conducted across Stages 1 and 2 of the evaluation, alongside a document review. An impact case study report was produced at Stage 2. As this provides insights into potentially sensitive government processes, the report is internal to DFID and DPME, but has been drawn upon in earlier synthesis reports as well as this final report (see Figure 3).

Component 4. Synthesis

This overview report brings together the findings from the full three years of evaluation outputs: the Stage 1 and 2 programme evaluations and Stage 3 country case studies, the literature review and impact case study, and the Stage 1 and 2 synthesis reports. It aims to draw generalisable conclusions about how and why different BCURE interventions have contributed to different patterns of outcomes in different contexts, producing an evidence-based set of refined CIMOs and a refined programme theory. Figure 3 presents a summary of the data from various evaluation components, illustrating how this has fed into this final report.

The synthesis process involved:

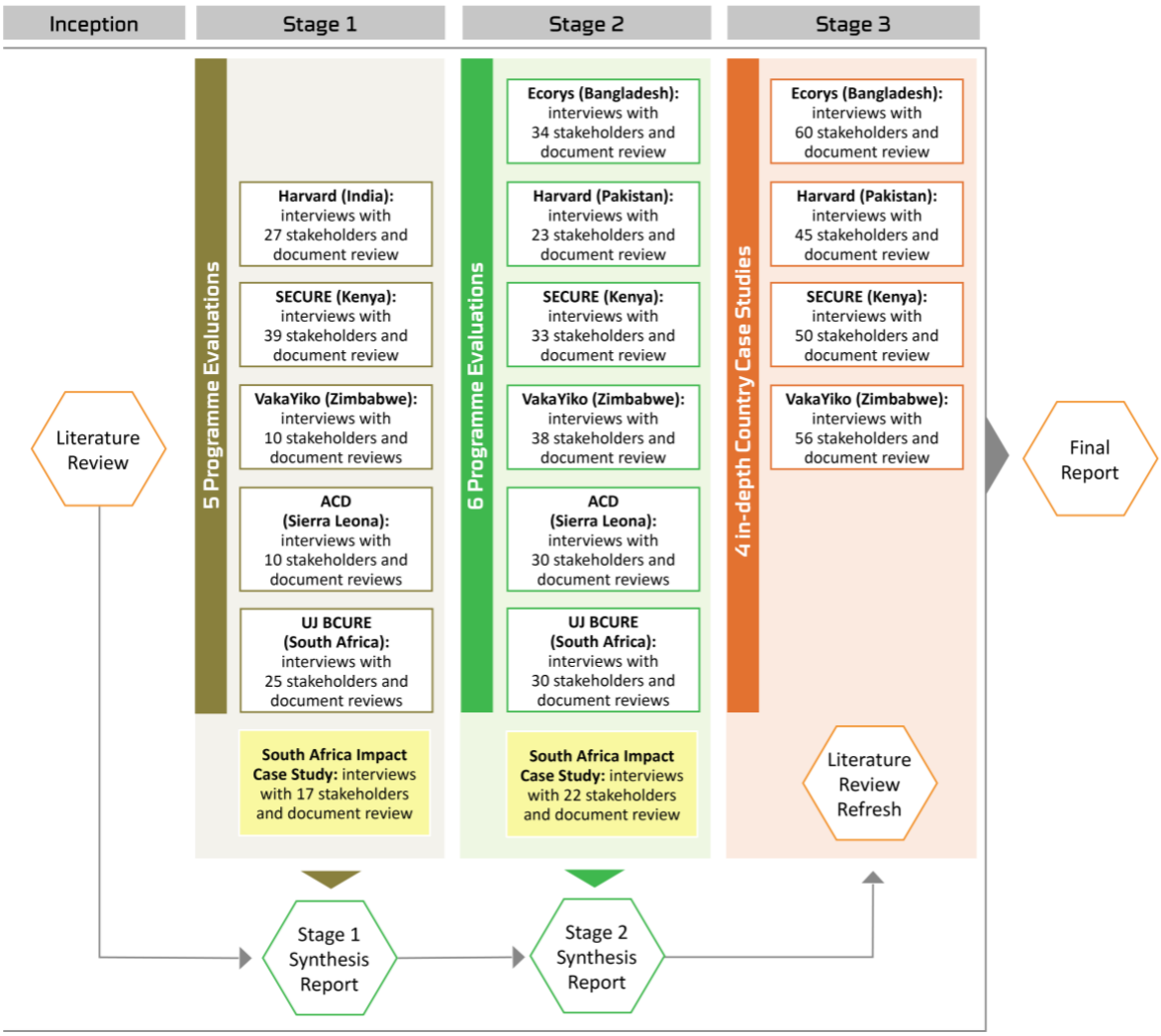
1. **Using a synthesis database to combine relevant evidence** from across the four Stage 3 case study reports about the outcomes achieved and not achieved (EQ 1), BCURE's contribution to these outcomes (EQ 2) and how and why particular outcomes were and were not achieved (EQ 3).
2. **Conducting a realist synthesis across the cases**, exploring how and why different BCURE interventions contributed to different patterns of outcomes in different contexts (EQ 3), in order to produce evidence-based set of refined CIMOs. This process applied realist synthesis techniques and additional insights from meta-ethnography in order to draw out meaning in a systematic way (described further in Annex 3.7). As well as the Stage 3 case studies, this process also drew on the Stage 1 and 2 synthesis reports, the impact case study report and the literature review.
3. **Checking and validating emerging conclusions**, through reviewing case study reports and where necessary interview data, to ensure that the evidence used to support, refute or refine the hypotheses underlying the findings was relevant and sufficiently rigorous to support the inferences made. The two lead researchers also cross-checked each other's analysis and conclusions, and shared drafts with other members of the core evaluation team to further validate and nuance findings.

During the second step a few broad and cross-cutting patterns emerged that appeared to explain incidences of success across the BCURE portfolio. These patterns were discussed within the evaluation team, and then systematically analysed by developing matrices that drew together relevant insights from across the sources, and applying the synthesis techniques described in Annex 3.7. This analysis suggested the importance of three broad 'ways of working' when seeking to build capacity for EIPM, described in Section 5.

Throughout the findings sections, specific insights on 'what worked, for whom, and why' have been drawn out in summary boxes. These represent 'empirical' CIMOs, which explain **outcomes** (O) from across the BCURE projects in terms of the **mechanisms** (M) that were (or were not) sparked by resources provided by BCURE, and the **context** (C) and **intervention** (I) factors that enabled or constrained the mechanisms. In the conclusions, these empirical formulations are brought up to the level of middle-range theory,[4] representing our final tested theory about what works to build capacity for EIPM, for whom, and in what circumstances.

As Figure 3 illustrates, the final report draws on a large amount of data collected over three years. The BCURE programme and evaluation are complex and multi-faceted, and comments on the Stage 2 synthesis report emphasised the importance of finding a clear way to present the findings in order to make them accessible to key audiences. It was therefore decided (in collaboration with the evaluation Steering Committee) not to structure the report according to the evaluation questions, but instead to use a simple overarching framework. The three impact pathways (discussed in Section 3.2) provide this framework, within which the evaluation unpacks what happened (EQ 1), BCURE's contribution (EQ 2) and how and why BCURE did or did not contribute (EQ 3).

Figure 3. Summary of data feeding into the final evaluation report



[4] This is theory that is "*detailed enough and 'close enough to the data' that testable hypotheses can be derived from it, but abstracted enough to apply to other situations as well*" (Wong et al., 2013).

Judging strength of evidence and extent of contribution

'Strength of evidence' relates to the internal validity of the evaluation findings. Our aim through the Stage 3 evaluation was to achieve a sufficient degree of confidence about the extent to which priority outcomes have occurred, BCURE's level of contribution to the outcomes and our theories (CIMOs) about how and why BCURE contributed or failed to contribute. Confidence in our conclusions about outcomes, contribution and CIMOs is underpinned by three broad considerations:

1. **The extent of triangulation** across stakeholders, participants/non-participants, and/or data sources.
2. **A consideration of the position, knowledge, analytical capacity, reflexivity, and potential biases of primary informants** – recognising that individuals are positioned in unique ways in relation to the programme, with different levels of knowledge and capacity / willingness to analyse and reflect, as well as different incentives that may lead to bias.
3. **A consideration of the broader context** – helping to ensure that explanations of change are grounded in an understanding of the political context and are not over-reliant on the explanations of programme participants.

These three considerations were used to develop a qualitative approach to assessing the strength of evidence (see Annex 3.8). This was not designed to be a rigid framework, but rather a way to ensure that evaluative judgements were made systematically and would be comparable across the case study and overview reports.

The overview report provides high-level comparative insights across the BCURE portfolio, basing the findings and judgements about the strength of evidence on findings from the Stage 3 country case studies and (for the two programmes that were not investigated at Stage 3) the Stage 2 programme evaluation reports. These reports, which are internal to DFID, discuss the evidence base underpinning the findings in detail, including through reference to specific interviews and data sources.

Stakeholder engagement throughout the evaluation

The BCURE evaluation has been designed and implemented in close collaboration with the DFID evaluation Steering Committee, through regular meetings and calls, as well as numerous internal approach papers which offered an opportunity for DFID to review and comment on emerging design choices. This regular engagement has facilitated annual revisions to the design in order to ensure the evaluation is meeting DFID's needs, particularly at Stage 3 where a substantial redesign was conducted (described above). The Steering Committee were also consulted on the selection of priority outcomes and CIMOs to test at Stage 3, based on the issues and questions most relevant to the design of future programmes.

BCURE partners have also been engaged at various points throughout the evaluation. Annual BCURE learning events offered an opportunity for the evaluation team to share emerging findings and interim analysis, with comments from partners fed into synthesis reports. In-country workshops with project partners provided an opportunity to hear the views of implementation teams and test CIMOs against their understanding of how and why change was (or was not) happening. At Stage 3, draft country ToCs were discussed with BCURE partners, and revised accordingly. Where possible during country visits, the evaluation leads also conducted debrief interviews or workshops with project staff, to share emerging findings at the end of the fieldwork, answer partner questions, and sense-check interpretations. Finally, draft programme evaluation and country case study reports were shared with partners to provide an opportunity for comments before the reports were finalised. These reports are internal, in order to protect the confidentiality of respondents and the relationships of BCURE partners with government stakeholders. However, synthesis reports and other publicly available evaluation products have been shared with interviewed stakeholders.

2.4 Limitations and lessons for future evaluations

The evaluation team was able to work freely and without interference, and there are no conflicts of interest to report.

The Stage 3 evaluation attempted to address a number of limitations identified at Stage 2. While this was successful to a large extent, certain issues proved difficult to address, giving rise to important lessons for future realist evaluations and evaluations of EIPM capacity development programmes.

The realist evaluation approach has been challenging to implement across the complex BCURE programme. In particular, we have faced challenges with 'breadth vs depth', including how to systematically prioritise outcomes and theories to assess within the limited time available for interviews. The Stage 1 and 2 evaluations generated a large number of theories (CIMOs) about how and why BCURE might be contributing to change at individual, interpersonal, organisational, institutional and policy levels. At Stage 2, it became clear that it was not possible to systematically test theories across the whole BCURE theory of change with the resources available for the evaluation. This was mitigated at Stage 3 through conducting a smaller number of more in-depth case studies, and prioritising a smaller number of outcomes and CIMOs for investigation. Developing country-level ToCs rather than relying on an overarching ToC helped identify case-specific outcomes and CIMOs that were less well-evidenced through earlier stages of the evaluation, and which were most important for achieving longer-term outcomes. This approach proved largely successful, and highlights the importance of realist evaluations prioritising the most interesting and important causal links in enough depth to draw useful insights, rather than trying to investigate everything. Case-specific theories, rather than (or as well as) a single overarching theory, can help facilitate this, through building an in-depth understanding of how and why a programme is expected to unfold in a specific case.

It has also been challenging to encompass complexity within the CIMO framework, including features of the macro political context and how they give rise to or inhibit mechanisms of change. There is a risk that CIMOs become overly linear explanations of how and why change happens ('this intervention feature, in this context, sparks that mechanism to lead to this outcome'). This was mitigated by presenting the final CIMOs in a more narrative way, which allowed the nuances and interconnections to be unpacked. The Stage 3 CIMOs also contain multiple features of context and multiple mechanisms, illustrating how these work together to lead to outcomes.[5]

In Stages 1 and 2 of the evaluation, it proved much easier to identify 'micro' features of context (e.g. around the characteristics of trainees) than 'macro' features (e.g. around the nature of government systems, the influence of power, politics and high-level incentives). This was mitigated through incorporating a specific PEA step in the Stage 3 methodology, which significantly enhanced the richness of the analysis. However, while the interviews provided a wealth of insights into the risks and opportunities that the context posed for EIPM and the programme's desired outcomes, it was not possible to gain insights into certain important issues likely to affect evidence use, including actual distribution of power and decision making, and some of the individual and organisational incentives that affect evidence use by senior government stakeholders. This suggests the importance of building in an explicit PEA lens from the outset in future realist evaluations working in government contexts, and considering how the evaluation design and team can be structured to best gain access to information on power distribution and incentives.

The evaluation has by necessity relied on interview data for evidence of outcomes, and there is a real risk of positive (confirmation) bias of respondents. With some exceptions, BCURE project monitoring systems were not set up to capture evidence of outcome-level change (including behaviour change and changes in policy processes or content). This has proved a major challenge for

[5] This follows the example of Leavy, Boydell and Mcdowell (2017).

the evaluation, suggesting the importance of ensuring future programmes build monitoring systems that go beyond measuring outputs such as self-reported increases in knowledge and skills. The evaluation explored the possibility of conducting large-scale surveys to capture insights from a broad cohort of participants, but given the high risk of low response rates this was not pursued. As a result, the evaluation has relied primarily on qualitative interviews with a select number of participants in order to provide evidence of longer-term outcomes. This carries a risk of confirmation bias, given the power dynamics of interviewing government stakeholders in low and middle-income countries. There is also the risk that participants may genuinely feel that the programme contributed to a positive outcome, when in fact other factors were more important – and this risk is heightened due to the complexity of the interventions, which makes it challenging to unpack contribution. We attempted to mitigate this at Stage 3 as follows:

1. **Triangulation:** We aimed for no more than 60% of the sample to consist of project participants and programme staff, with the remainder consisting of knowledgeable non-BCURE participants. Increasing the number of non-participant interviews helped to triangulate insights from project participants with the perspectives of individuals with less stake in the programme and potentially less incentive to tell the evaluators what they felt we wanted to hear.
2. **Conducting a more in-depth investigation into priority outcomes, and identifying and testing non-BCURE influences of change.** Focusing on a small number of priority outcomes enabled us to interrogate stakeholder testimony in more depth, helping us gain more detailed insights into what had happened and what had enabled or inhibited change. Our PEA exercise provided insights into country and sectoral contextual dynamics, helping ensure that explanations of change were grounded in an understanding of the political context, were not over-reliant on the explanations of programme participants, and were fair to programmes working in challenging settings. This helped to guard against over-attributing change to BCURE, as well as contextualising shortfalls in programme achievements.
3. **Dedicating more resources to finding monitoring and other documentary sources in order to triangulate interview data.** This included policy documents or government guidance that would help us validate stakeholder testimony about improvements in evidence access, appraisal and use. While this had some success and most case studies were able to view at least some documentation, there were ongoing challenges in accessing this data as the majority of stakeholders were unable to share internal government documents. This challenge was somewhat mitigated through interviewing a wide range of participants, and where possible their colleagues and managers, to triangulate insights.

It proved very challenging to secure interviews with government officials across all four settings – both participants and non-participants – particularly in Bangladesh and Pakistan. Challenges in securing and conducting interviews was a result of high workloads, adverse weather, security concerns (in Pakistan), and also the fact that most BCURE projects had largely finished activities in both settings, providing limited incentive for participants to volunteer their time. These challenges were mitigated through dedicating substantial efforts to contacting and following up with respondents, and through extending the length of country visits; however, in a number of cases the interviews were very short and it was only possible to explore a small number of outcomes and theories. This is reflected in the depth of analysis and strength of conclusions drawn in the country case study reports, and subsequently this overview report.

It has been challenging to ensure consistency of data collection and analysis across a diverse programme team. Time and budget constraints meant it was challenging to train the team comprehensively on the evolving programme theory, the principles of conducting realist interviews, and the approach to analysing data in a realist way. This created issues with ensuring a consistent approach to testing CIMOs and analysing interview data across the cases. At all three evaluation stages, we have attempted to mitigate this through a team workshop prior to data collection, involving a full introduction to the programme theory and basic training on realist interviewing and analysis. Programme leads provided training in-country to national evaluators prior to data collection, and additional analysis was conducted at synthesis stage by the team leader and methodological lead to capture insights that may have been missed during the case study analysis. At Stage 3, we also revised the team structure so that country visits were conducted by two core team members rather than one, which helped improve consistency across the cases. However, our major reflection is that realist evaluations require a different approach to team structuring and capacity building. Realist interviews and analysis require team members to have an in-depth understanding of not only the methodological approach, but the theory that the evaluation is trying to test. In order to ensure consistency and understanding, a realist evaluation requires a more participatory model, which involves in-depth and ongoing engagement and capacity building.

Finally, the evaluation draws on evidence from only six of the 12 BCURE countries, and the short time frame of the programme limits the potential to record longer-term results. The evaluation is limited in what it can say about how BCURE worked across all of its settings, because its focus on six country case studies means it has not captured the full range of outcomes across the whole portfolio. Given the country-level focus, it also does not capture outcomes from the international and regional networking components that were part of various projects.

Finally, the programme was relatively short given its aim to generate systemic change in government settings – the shortest programme, in Bangladesh, had only two years of implementation time. This has limited the ability of the evaluation to identify longer-term results.

3. What was BCURE's approach to improving capacity for evidence use?

3.1 The BCURE projects

BCURE was made up of six projects, implemented by different partners, and conducted over four years from 2013–17 (see Table 1). The programme was implemented in 12 core focal countries in Africa and Asia, as well as a number of additional countries reached through cross-country networking and small-grant initiatives. The projects were implemented in countries with low or mixed use of evidence in government decision-making, affected by various political economy dynamics in the national and sectoral contexts. They all focused on building skills, networks and organisational systems for EIPM, through a range of interventions designed and combined in different ways. Projects ranged in scope and scale, from working in single ministries to working across whole government systems. Implementing partners developed close operational relationships with government and other institutional entities to implement the interventions collaboratively, a crucial aspect of the BCURE approach.

Table 1. Summary of BCURE projects

Project name	Lead implementing partner	Main implementation countries *(case study country in bold)*
Africa Cabinet Decision-Making (ACD) Programme	Adam Smith International (ASI) and the Africa Cabinet Government Network (ACGN)	**Sierra Leone**, South Sudan, Liberia
Building Capacity for the Use of Research Evidence	Ecorys	**Bangladesh**
Data and Evidence for Smart Policy Design	Harvard	**Pakistan,** India, Nepal
SECURE Health	African Institute for Development Policy (AFIDEP)	**Kenya**, Malawi
UJ-BCURE	University of Johannesburg (UJ)	**South Africa**, Malawi
VakaYiko	VakaYiko Consortium, led by INASP	**Zimbabwe**, South Africa, Ghana, Uganda

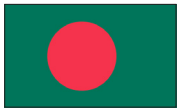
BCURE was a pilot, and was not commissioned by DFID as a cohesive programme but rather as a portfolio of different projects with their own perspectives on EIPM. While the overall long-term objective was always to increase the use of evidence in policymaking, this was interpreted in various ways by the different partners. For the most part, projects were initially designed with a fairly technocratic approach with limited political economy analysis, in line with DFID's original terms of reference, and reflected in a contracting model that involved up-front milestones and limited flexibility. As BCURE progressed, there was more critical questioning of the basic premise of BCURE amongst DFID and the partners, and the framing became more politically informed over time. Partners had to adapt their programmes to more responsive ways of working with governments within the existing contract frameworks.

The project descriptions on the next few pages focus on the six countries visited as part of the evaluation. As described in Section 2.3, the Stage 3 evaluation focused on four BCURE countries: Bangladesh, Kenya, Pakistan and Zimbabwe. Insights from the Stage 1 and 2 studies in South Africa

and Sierra Leone are also drawn upon in this report. To avoid confusion, throughout the report we refer to the case study country rather than the project or implementation partner.[6]

As noted in the introduction and Section 2.3 above, the report only includes insights from the six countries in bold in Table 1, as the decision was made to focus in-depth on a smaller number of contexts in order to tell a more nuanced story about how and why the programme worked and did not work in specific settings.

[6] While two BCURE projects operated in South Africa (UJ-BCURE and VakaYiko), this report only includes insights from UJ-BCURE's work in the country. It also includes insights from the South Africa impact case study – this is clearly differentiated in the narrative.



Bangladesh

Context

Bangladesh is a parliamentary democracy, although genuine representational politics remains aspirational, and the country scores low on indicators of governance and has a record of high corruption. However, public sector reform to improve government effectiveness has seen progress through programmes supported by World Bank, UNDP and DFID, which creates opportunities for promoting better evidence-informed policymaking.

As one of the most powerful regulatory ministries, Cabinet Division has gradually taken on the role of coordinating and championing reform. Nevertheless, major challenges to improving the use of evidence in decision making remain, including few incentives for policymakers at senior levels to consider evidence, with patronage exerting a strong influence on decisions and many policy issues remaining highly contested.

Ongoing capacity shortcomings in the civil service include inadequately trained personnel, weak or overstretched systems, basic lack of access to data/evidence by various stakeholders, and poor links or tense relations between government and non-governmental actors.

BCURE project

In Bangladesh, BCURE was implemented by Ecorys, in partnership with the Dhaka-based Policy Research Unit (PRI). This programme had a shorter timeframe than the others due to delays in the design phase, commencing in 2015 and running for two years. The programme worked with Cabinet Division to develop EIPM guidelines and pilot them in three line ministries: Commerce (MoC), Health and Family Welfare (MoHFW) and Environment and Forests (MoEF).

An EIPM training course linked to the guidelines was piloted in the three ministries, designed to impart technical skills to access, appraise and apply evidence in policy formulation, predicted to reach 400 civil servants by the end of the programme in November 2017. BCURE also provided on-the-job consultancy support to six ‘pilot policies’ in the three ministries, helping policymakers apply the guidelines to specific policy development processes.

Finally, BCURE worked with national training institutes to build the capacity of trainers and embed EIPM training in existing civil service training courses. Cutting across all of these activities, the programme engaged senior stakeholders in Cabinet Division and line ministries to build interest in and buy-in for EIPM, including through learning events and an exchange visit to Indonesia.

Kenya

Context

Kenya is a functioning multi-party democracy, although Kenya is rated as only 'partly free' by Freedom House due to political and bureaucratic corruption and regular state overreach.

Challenges to improving the use of evidence in decision making include politicised ethnic allegiances and patronage networks which exert a strong influence over decision making in both government and Parliament, providing minimal incentives for policymakers to consider evidence at senior levels.

However, several key drivers are creating opportunities to enhance the use evidence. First, there are political pressures on government to improve service delivery, especially in the health sector, which, alongside high levels of international donor investments, are creating incentives to use evidence. Second, the civil service is seen as generally impartial and merit-based, and is slowly adopting performance contracting and monitoring, pushing forward a fledgling culture of results and evidence. Third, since 2013, Kenya has been transitioning to new governance and administrative roles mandated by the 2010 Constitution, which has stimulated demands for data and research in national and county governments, as well as in Parliament, which has decision-making authority in the new constitution.

BCURE project

In Kenya, BCURE was implemented by the SECURE Health consortium, led by the African Institute for Development Policy (AFIDEP). The programme worked in partnership with the Research and Development (R&D) unit at the Ministry of Health (MoH), and the Parliamentary Research Services (PRS) unit in Parliament.

It delivered EIPM training and coaching support to 45 officials in the two organisations, including a secondment to the UK Parliamentary Office of Science and Technology (UK POST) for two researchers. The programme also conduced eight science policy cafes and other events bringing together policymakers and researchers to discuss evidence access and use, provided technical support to develop EIPM guidelines, and supported policy processes – including a national Research for Health policy.

Finally, the programme conducted a number of presentations on research use at high-level meetings in Kenya and Malawi, including at regional meetings.

More information

Pakistan

Context

Pakistan is a federal parliamentary republic, with constitutional democratic elections regularly held for national and provincial governments, although the transfer of power between administrations is often turbulent.

The country faces major challenges in terms of security and law enforcement, economic reform and infrastructure development. Pakistan has a large civil service, and there are national data systems and research; however, evidence use is limited in government due to a range of challenges including the ongoing politicisation of the civil service.

Positive opportunities to improve the use of evidence include a wide-ranging reform process led by the National Commission's Civil Service Reform Framework and a growing momentum around the use of technology and digitisation to improve government information systems.

BCURE project

In Pakistan, BCURE was implemented by Harvard University in partnership with the Centre for Economic Research. The main activity was a large-scale training programme that reached 1,780 civil servants in Pakistan, aiming to equip policymakers with the knowledge and skills to use evidence more effectively for decision making.

The programme also conducted three policy dialogues seeking to promote learning between policymakers and technical teams to provide solutions to EIPM challenges.

Finally, the programme conducted a number of pilot projects, aiming to develop practical tools to facilitate EIPM through improving the presentation and visualisation of public sector data.

More information

Sierra Leone

Context

Sierra Leone is a presidential representative democratic republic. The president exercises executive power and parliamentary legislative power, within a multi-party system, although challenges exist in terms of government capacity and lack of trust in government institutions. The cabinet system is functioning but there is limited development of the system as a legitimate mechanism for balancing competition for resources and coordinating across ministries.

Positive opportunities to improve the use of evidence include an acknowledged need for greater use of evidence to support implementation of decisions. A major post-Ebola recovery programme was launched in 2015 involving more than six key ministries of government.

One impact of Ebola is the reprioritisation of the government of Sierra Leone's development plan – the Agenda for Prosperity – towards health, education, water, energy, agriculture, social protection, private sector development and governance, creating further opportunities to promote evidence use.

BCURE project

In Sierra Leone, BCURE was implemented by Adam Smith International (ASI), in partnership with the Africa Cabinet Government Network. The programme provided ongoing advisory and technical support to the Cabinet Secretariat through a national advisor embedded in the Cabinet Secretariat, to develop revised and customised manuals for Cabinet procedures, and templates that required consideration of evidence in proposals to Cabinet.

International training, roundtables and on-the-job support were provided to the Cabinet Secretary and other staff in Cabinet, to build capacity to undertake policy analysis and oversee the implementation of the new procedures.

A Cabinet focal person (CFP) network was also established in line ministries, with training and on-the-job support provided to the new CFPs.

More information

South Africa

Context

South Africa is a relatively young constitutional democracy, regarded globally as a proponent of human rights and a leader on the African continent since the end of apartheid in 1994. Elections are regular and generally considered free and fair, while Parliament has oversight over the Executive. South Africa is a relatively 'high capacity' environment, which has provided opportunities to embrace and extend EIPM practices in ways that have not been possible in the other countries. However, the use of evidence in decision making in the South African government is mixed.

Strong ideological affiliations on all sides of the political spectrum tend to determine policy priorities and patronage relationships exert a strong influence on decisions, which skews incentives for policymakers at senior levels to consider evidence. However, the EIPM agenda has had an institutional advocate since establishment of the Department for Planning, Monitoring and Evaluation (DPME), in the Office of the Presidency in 2009. The DPME focuses on increasing monitoring and evaluation systems in the South African public sector and has, since 2014, also incorporated the planning function under its mandate, offering focused EIPM evaluation capacity building to senior managers, as well as support to Parliament.

The positive influence of global organisations as well as regional forums such as the Southern Africa Development Community and the African Union, supports the use of evidence, as they are seen to demand evidence as the basis for programming decisions.

BCURE project

In South Africa, BCURE was implemented by the University of Johannesburg. The programme worked with the DPME as well as civil servants in a number of line ministries: initially the Department of Science and Technology and the Department of Basic Education.

The programme conducted EIPM workshops with civil servants and stakeholders from research and civil society organisations, primarily aiming to raise awareness of and build buy-in for EIPM, as well as facilitate dialogue and network building. One-to-one mentorships (13 as of July 2016, several months before programme completion) were facilitated between South African sectoral and knowledge experts and junior, mid and senior-level civil servants in the focus ministries, to help improve individual skills and confidence in applying evidence in their day-to-day work. Later in the programme BCURE conducted a small number of 'team mentorships', providing technical support to help facilitate evidence use within an organisational process – including working with the DPME to facilitate the production of an evidence map.

Finally, the programme facilitated and supported the Africa Evidence Network, overseeing a membership drive, supporting the website, and coordinating an international Evidence conference.

More information

Zimbabwe

Context

Zimbabwe is a presidential republic, dominated by President Robert Mugabe and his Zimbabwe African National Union – Patriotic Front (ZANU-PF) since independence in 1980. Mugabe resigned in November 2017, shortly after the completion of the BCURE programme and just before the finalisation of this report. In a government usually described as 'authoritarian', there is regular and overt intrusion of the security sector into politics, making the Zimbabwean political landscape deeply contested and fragmented, which disrupts government operations and limits the space for evidence to inform decision making. Zimbabwe also faces severe ongoing economic challenges and corruption issues. Severe resource constraints limit government activities, whose programmes are in large part funded by international donors. Resource constraints also contribute to an under-capacitated civil service as government has been unable to invest in training. Despite the many challenges, several stakeholders felt there is growing interest in the concept of EIPM within government, in part driven by the sense of an urgent need to find solutions to the nation's problems. There may be pockets of opportunity for EIPM in sectors such as health, education and the environment, where there is more receptiveness to evidence because there are fewer political implications. Further, there are a number of ongoing reforms that offer opportunities to promote EIPM, including an Integrated Results-Based Management system (2005) and a National Monitoring and Evaluation Policy, and high levels of donor investment in programming also drives evidence use through the need to justify expenditure.

BCURE project

In Zimbabwe, BCURE was implemented by the VakaYiko consortium, led by INASP and implemented in Zimbabwe through ZeipNET. The programme trained 49 civil servants across the Ministry of Industry, the Ministry of Youth, Economic Empowerment and Indigenisation (MoYIEE), and Parliament, aiming to build technical and soft skills for EIPM among research and policy staff at a junior to mid-level.

The training was followed up by a 'mentorship and exchange' programme, which aimed to build on the training through addressing organisational-level barriers faced by trainees, and providing parliamentary trainees with the opportunity to learn from peers in Ghana and Uganda through an exchange programme.

VakaYiko also conducted six policy dialogue and four knowledge cafe events, aiming to facilitate evidence-informed discussion around particular policy issues, raise awareness about and demand for EIPM, and facilitate collaboration and evidence-informed dialogue. In total, approximately 370 people directly took part in these events, including a range of stakeholders from government, civil society, academia, the media and the general public.

More information

3.2 How and why was BCURE expected to lead to change?

This section presents the evaluation's theory about how and why BCURE was expected to lead to change, and describes how this theory evolved over the course of the evaluation.

Stage 1 BCURE ToC: Unpacking 'capacity development' to create a unifying framework

The six BCURE projects were highly diverse, taking different approaches to enhancing skills and systems for evidence use, in complex government contexts. As BCURE did not have a portfolio-level ToC, the evaluation developed an initial Common Theory of Change (CToC) through synthesising the original project ToCs and incorporating key insights from the literature review. The Stage 1 CToC unpacked capacity development into four levels of capacity change,[7] which helped to bring the diverse BCURE approaches into a unifying framework for the evaluation. The four levels – individual, interpersonal, organisational and institutional – conveyed the concept of capacity development as multidimensional, and capacity as a function of different factors and processes working together and reinforcing each other at different levels. The BCURE ToC at Stage 1 stated that multidimensional change across these four domains would contribute to routine change in the use of evidence by government, in turn contributing to improved *quality of policy development processes*, as the overall impact.

Stage 2 ToC: Unpacking 'evidence use' and EIPM as a system, and defining 'policy quality'

The Stage 2 evaluation confirmed our theory that changes emerging at different levels (e.g. individual skills and organisational systems) seemed to reinforce each other, and that changes at different levels were required to make progress towards improvements in the quality of policy products and processes. This led us to revise the CToC, representing capacity as a set of interlocking domains, with entry points at individual, interpersonal, organisational and institutional levels.[8] However, changes at the interpersonal level were difficult to find conclusive evidence for at Stage 2; and changes in the institutional domain were both beyond the scope of the evaluation to investigate and a relatively minor part of most project activities. In response to this (and the challenges of breadth and depth noted in the methodology), the next iteration of the ToC focused on a prioritised set of theories about how and why individual and organisational change supports improvements in evidence use and changes in policy quality. See Annex 4 for the theories prioritised for testing at Stage 3.

The BCURE literature review identified an inherent tension between more rational, technical perspectives on 'evidence use' and 'policy quality', and perspectives that view EIPM more as a complex system, infused with power and shaped by the political economy of the country context, leading us to develop a dynamic definition of 'policy quality' at the end of Stage 2. Politically informed perspectives on evidence use challenged the basic premise of the BCURE programme that improved and more routine use of evidence leads to better quality policy development, especially given the complex realities of government settings in the BCURE countries. This insight made it difficult to reconcile 'rational' interpretations of 'policy quality' that had informed many of the BCURE partners' approaches (see Box 2). The conclusions of the literature review suggested the value of engaging critically with this concept over the course of the evaluation.

[7] There are many definitions used in the literature to describe levels of capacity change. We adapted DFID's definitions from the 2010 'How to Note on Capacity Building in Research' (DFID, 2010). This document uses 'institutional' to denote 'changes in the rules of the game'. Other readers may interpret 'institutional' to mean 'systemic' or 'environmental' change. We opted to consider the government system as falling within a broadly conceived 'organisational change' category because organisations within the government system are bound by common, cross-cutting rules, incentives and procedures. This means that 'institutional' change then encompasses all non-governmental influences within the wider environment. However, we recognise that the boundaries between the levels of change are fuzzy and dynamic.

[8] The diagram and full narrative of the Stage 2 programme theory, including the CIMOs, can be found in Annex 4.

Therefore, a working definition of 'policy quality' was developed for the Stage 2 ToC and evaluation, to enable the evaluation to take into account dynamic aspects of evidence use (and misuse) in the investigation of BCURE's overall outcome. Our definition proposed that, for evidence use to promote critical thinking, a decision-making process needs to be *transparent* about the limitations of evidence by engaging explicitly with diverse perspectives and values and multiple types of evidence, and it should be transparent about the extent of evidence and its quality. In this way, productive debate and discussion on the issues raised by evidence can be encouraged and evidence is likely to have a demonstrable influence on the decisions made, whether conceptual or instrumental. However, occasional uses of evidence are not enough to achieve the impact. A key part of BCURE's intended impact was for evidence use to become *embedded* in decision-making routines, supported by organisational systems and incentives to use evidence.

At the end of Stage 2, it became clear that the evaluation needed to gain a deeper understanding of key political economy dynamics in order to explain BCURE's emerging outcomes. Therefore at Stage 3 we developed country-specific ToCs, drawing on the Stage 2 CToC and insights from the first two years of the evaluation, to provide a more concrete and contextualised framework for the country case studies. This allowed us to explore the critical political economy dynamics affecting observed and hypothesised causal links, as described in Section 2.3.

The definitions of 'policy quality' and overall impact were also revised, to enable an assessment of progress towards impact in the final evaluation. We unpacked our working definition of 'policy quality' from Stage 2 into four priority outcomes and an impact statement, to reflect *embedded, transparent, conceptual and instrumental* uses of evidence (see Table 2). These concepts linked to emerging DFID thinking on measuring the VfM of evidence-into-use interventions, and also linked to key insights in the EIPM literature (see Box 2). In this framing, the evaluation recognised that evidence use may be *strategic, tactical* or *political*, or may represent *misuse*, all of which would fall short of BCURE's anticipated impact.

Box 2: Insights from the literature: Understanding 'evidence use' in policy processes

The BCURE literature review highlighted the different ways that evidence is used in policy design, decision making and implementation. Weiss (1972, 1980, 1982) emphasised that **instrumental** use of evidence, where specific evidence directly shapes policy choices, is only one way – and is often quite rare. More common is where evidence generates a slow 'enlightenment' as concepts and theories from research gradually percolate through society, *'coming to shape the way in which people think about social issues'*. This was labelled **conceptual** use of evidence by Nutley et al. (2007). However, evidence may just as frequently be used to justify or refine a position that has already been reached, which can be thought of as **strategic**, **tactical** or **political** use. There are also examples of unambiguous **misuse**, when poor quality findings are used, or tactical use of evidence intentionally justifies a bad practice (Nutley et al., 2007). Finally, there are examples of **over-use**, where a set of findings may become a new fad and be applied uncritically and wholesale.

Several sources in the literature review emphasise that evidence itself is not a neutral product – first because it reflects pre-existing views, values and beliefs of researchers and commissioners involved in producing it; and second because it rarely points to an obviously optimal solution, so contestation over its meaning is inevitable (see for example, du Toit, 2012; Waldman, 2014). This suggests the importance of considering the nature of the decision-making process, and how different evidence sources and stakeholder perspectives are consulted and interpreted.

Table 2. Definition of 'policy quality' used in the BCURE evaluation

Dimension a	Government officials routinely consider a range of evidence and the quality of evidence when developing policy products (embedded use of evidence)
Dimension b	Appropriate policy development processes engage with evidence from diverse stakeholders and multiple perspectives (transparent use of evidence)
Dimension c	Routine evidence use is facilitated by evidence tools, which allow officials to access, identify and critically appraise the evidence base and apply it to decisions, being transparent about the evidence base behind decisions (transparent use of evidence)
Dimension d	Routine evidence use is reinforced, incentivised and monitored through processes and standards, supported by senior managers (embedded use of evidence)
Impact	Together, a–d are expected to contribute to conceptual and in some cases instrumental uses of evidence…ultimately leading to evidence-informed decisions being implemented.

Stage 3 final report: identifying impact pathways towards improved use of evidence

The Stage 3 synthesis process compared the contextualised country case study ToCs to identify how the tested CIMOs had played out in the different countries. This highlighted three main routes towards EIPM taken by BCURE partners, at different levels of the government.[9] We have termed these 'impact pathways':

- **Impact pathway 1:** Support to a single ministry or unit
- **Impact pathway 2:** Working at a government-wide scale
- **Impact pathway 3:** Support to parliament

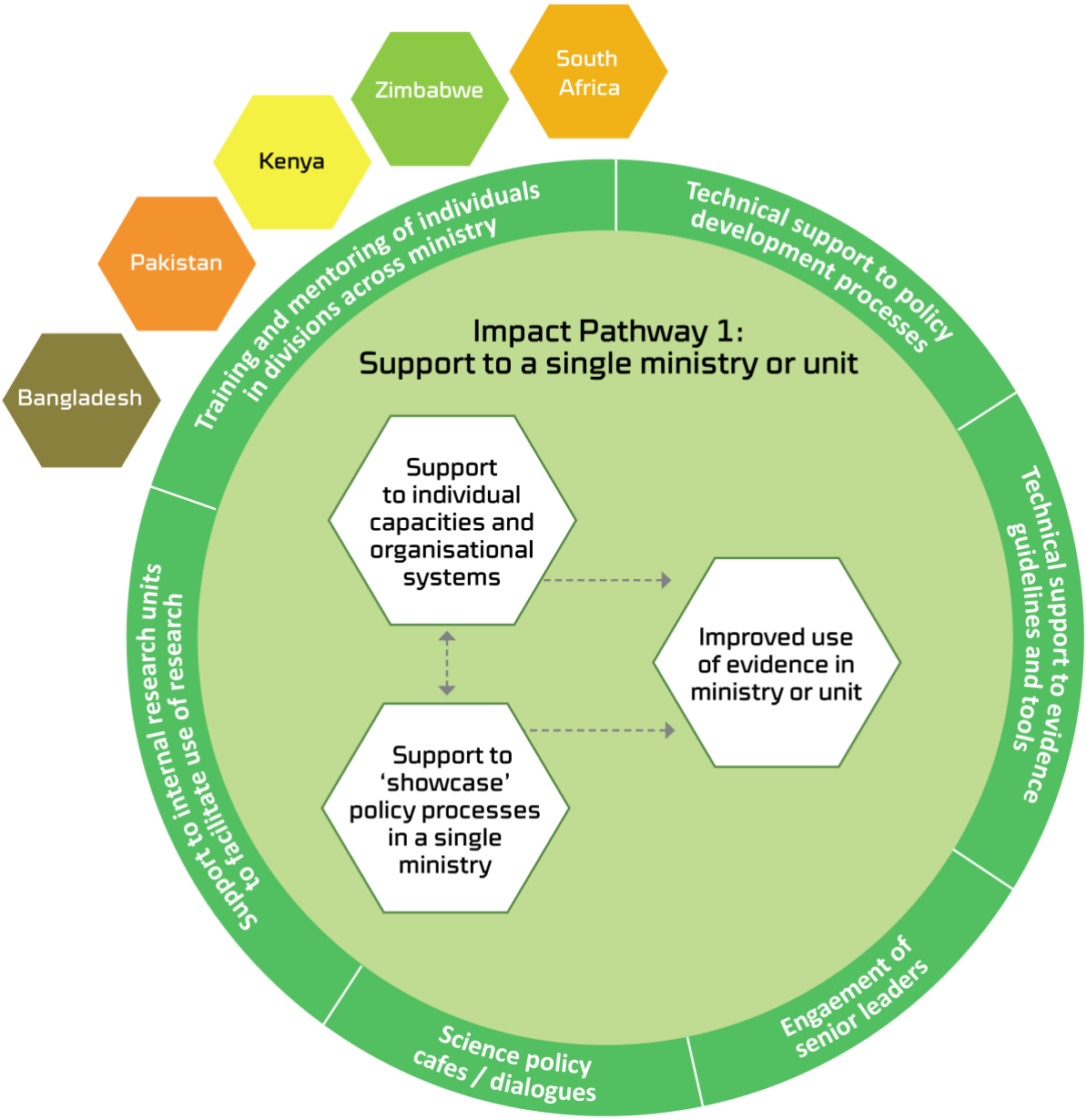
The impact pathways are archetypal programme theories (Funnell and Rogers, 2011), presenting a sequence of outcomes, from short-term to long-term, towards the desired impact, with evidenced causal explanations of how and why change has come about through BCURE. They are not mutually exclusive – most projects worked across two or more. These three impact pathways take the place of an overarching, portfolio-level ToC, providing a rich, context-specific explanation of how and why capacity support can promote EIPM through entry points at different levels.

The impact pathways focus on government and parliament, and do not encompass the important civil society and citizen engagement dimensions of EIPM. This is partly because few BCURE projects set out to engage citizens directly, and partly because the evaluation was only able to gather limited evidence on the BCURE activities that did aim to bring in civil society to the evidence process, for example, policy dialogues and science cafes. Spotlight 3 discusses these activities in more detail. The pathways also make limited reference to the international and regional networking components of BCURE projects (for example UJ-BCURE's support to the Africa Evidence Network), for the reasons discussed above (see p.24).

[9] The only project that did not fit neatly into the impact pathway classification was UJ-BCURE in South Africa. Here, the approach used-individual level entry points through providing one-to-one mentoring and workshops to government staff in multiple departments, rather than seeking to catalyse change within a single ministry or attempting to promote cross-government reforms. One element of the support provided – assisting the DPME to produce an 'evidence map' – is considered in Section 6, and further insights are available in the Stage 2 synthesis report (link), including around the one-to-one mentoring approach which was unique in BCURE.

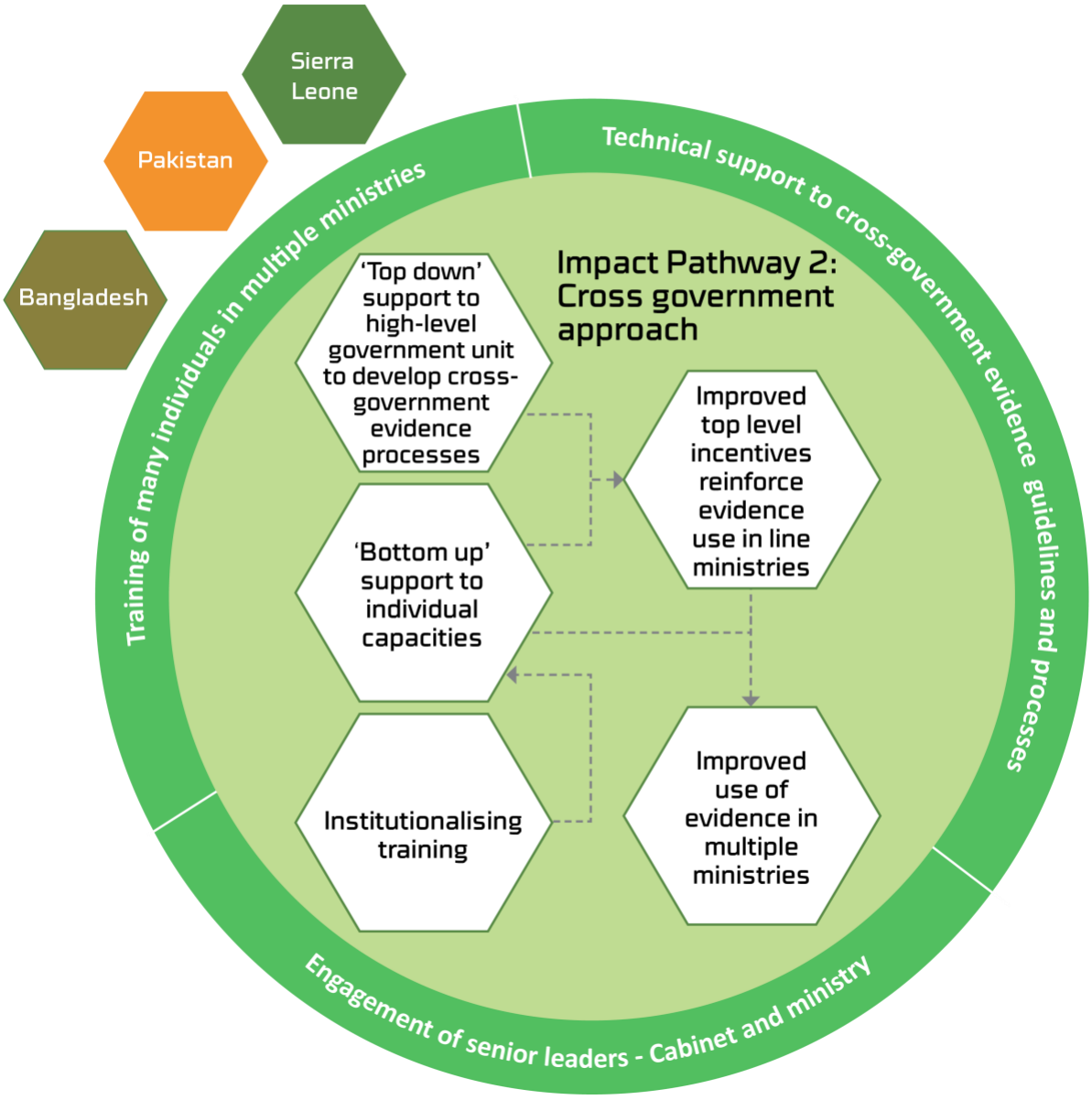
Impact Pathway 1: Support to a single ministry or unit

The single ministry pathway (see Section 5) incorporates two approaches: 'training-plus', and technical support to pilot policy processes or EIPM tools. In the 'training-plus' approach, training on EIPM was given to technical officers responsible for policy formulation, who were then provided with follow-up support, to help sustain the application of new EIPM skills. Organisational support was also given to tools and guidelines that were intended to be adopted by ministries in order to facilitate and incentivise individuals to use evidence more routinely and more skilfully. In the 'support to pilot policies' approach, some projects provided technical support at an organisational level to trial systematic, evidence-informed approaches to policy development within the ministry, providing EIPM trainees with opportunities to apply their skills, and producing evidence-informed policy proposals. Other projects provided technical support to develop data and evidence tools that aimed to showcase the value of evidence for decision making, with the hope that these would be adopted or replicated by government partners to help embed evidence use in the ministry or unit.



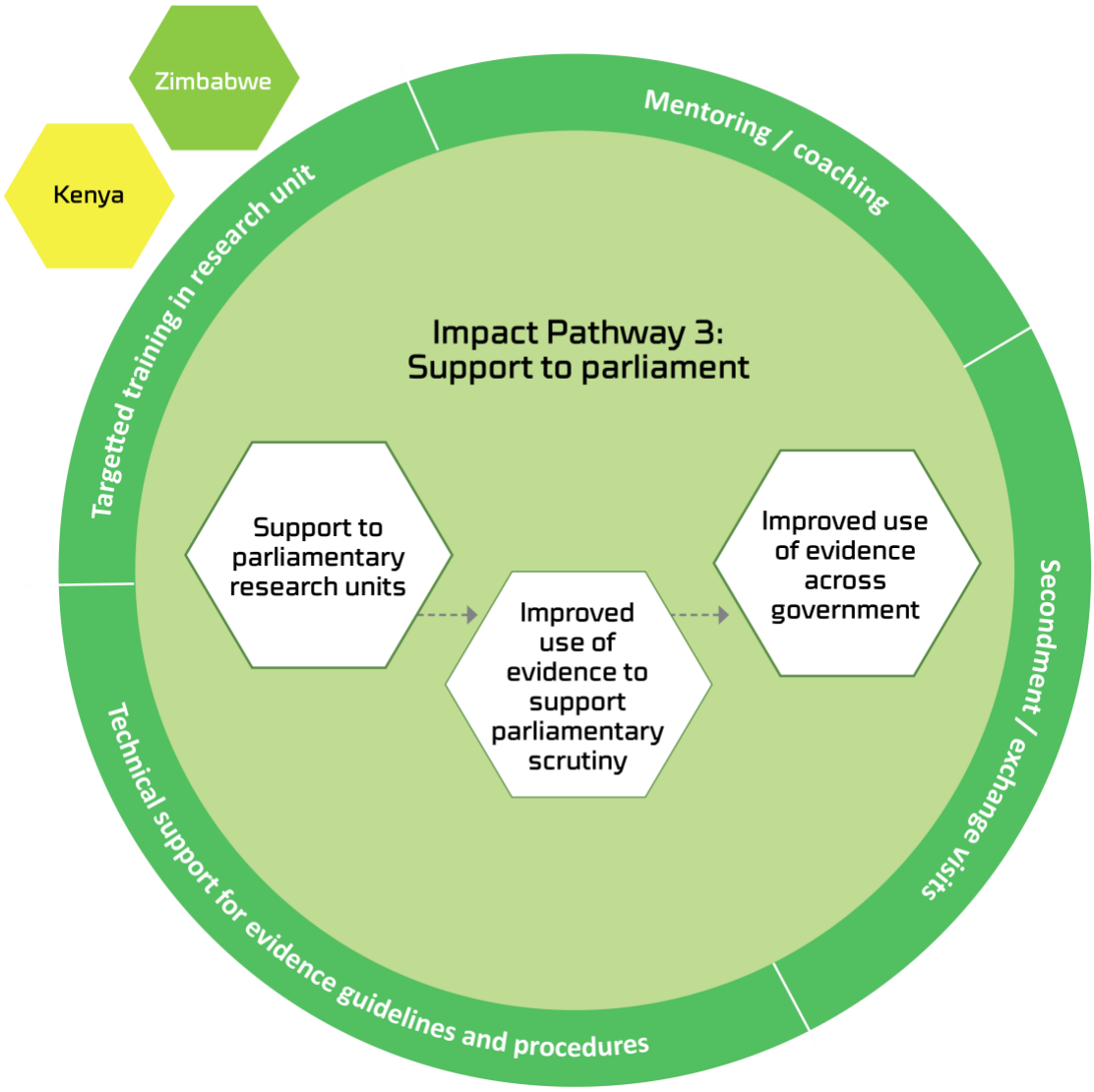
Impact pathway 2: Cross-government approach

The cross-government pathway (see Section 6) incorporates three approaches: 'top down', 'bottom up' and 'institutionalising training' to promote EIPM. In the 'top down' approach, BCURE-supported activities such as working with Cabinet to develop and roll out EIPM guidelines and procedures, often aiming to establish common cross-government standards for EIPM, facilitating various learning events, exchanges and training with senior government stakeholders to build high-level buy-in for EIPM, and supporting (mainly senior) officials in line ministries to develop evidence-informed policy processes, systems and structures in their ministries. In the 'bottom up' approach, BCURE programmes developed EIPM training courses and delivered them to large numbers of civil servants. The third approach involved BCURE working to institutionalise EIPM training, through embedding EIPM curricula within national public sector training institutes. In all three projects working across government, at least two of the three approaches were adopted simultaneously.



Impact pathway 3: Support to parliament

This setting poses a different set of issues and challenges – as parliaments do not make policy, but can play an important role in interrogating it and holding line ministries to account. The parliament impact pathway (see Section 7) incorporates training research staff within a parliamentary research unit (as an entry point to parliaments), combined with follow-up individual and organisational support to strengthen parliamentary use of evidence in oversight and scrutiny functions. The two BCURE projects taking this approach promoted EIPM by focusing on individual and organisational change in parallel, as in the single ministry pathway. Training on EIPM was provided to a cohort, or all, research officers responsible for supporting MPs and committees with impartial and accurate analysis of policies and bills, and with evidence for committee enquiries. Trainees were then provided with follow-up support to help sustain the application of new EIPM skills, as well as offered learning exchange opportunities with other parliaments to further enhance learning about EIPM. BCURE also provided ongoing technical support to help trained researchers cascade skills to non-trainees and develop procedures, ways of working, tools and guidelines that were intended to be adopted by Parliament in order to facilitate researchers, and sometimes MPs themselves, to use evidence more routinely and more skilfully.



Spotlight 1. BCURE's approach to gender, equity and social inclusion

The BCURE programmes took a fairly basic approach to integrating gender into the design and implementation of interventions, which involved collecting gender-disaggregated data on participation and aiming for equitable participation in activities. However, most programmes missed the opportunity to integrate gender perspectives as a key aspect of more effective policy development and analysis. The integration of gender was relatively piecemeal in the BCURE programmes. For example, the BCURE programme in Sierra Leone integrated a criterion on evidence of gender differentiation of impacts into its policy proposal template for line ministries, while in Zimbabwe BCURE incorporated gender issues within its EIPM training curriculum and knowledge cafes. BCURE in Pakistan also included two policy dialogues on women's economic engagement, which was a challenging topic in a country with high levels of gender inequality. Other programmes collected gender-disaggregated data on participation in activities, and aimed for a gender balance among participants where possible, but did not go beyond this. BCURE in Pakistan gathered data on gender differences in learning outcomes from the blended learning training, although the sample size prevented meaningful differences to be noted. None of the programmes gathered gender-disaggregated data on behaviour change or organisational outcomes.

Gender and inclusion considerations could also have informed the BCURE interventions, a lot of which were premised on individual civil servants being able to introduce change in their settings to promote the use of evidence, but no consideration was made of how gender and social inequalities might constrain individuals' power to influence change. For example, in most BCURE contexts, women officials were likely to experience more difficulties in influencing change than their male counterparts because of gender inequalities in male-dominated civil service settings. The challenges for women were perhaps most evident in Pakistan, where women have been systematically underrepresented in the civil service, and were only in 2006 guaranteed a certain amount of positions within the higher rungs of the civil service as part of an affirmative action plan. This meant that only 17% of the civil servants in the higher grades targeted by BCURE training in Pakistan were women.

Across all the BCURE projects, there was an opportunity to provide more tailored support to build women's leadership and influence, and integrate gender and inclusion perspectives in their EIPM training and support to policy processes. For example, this might have involved emphasising the importance of establishing a gender perspective from the outset when defining a policy problem, and ensuring as a fundamental principle that evidence and data is disaggregated by gender, age and socio-economic background to assess how women and men are differently affected by issues and the impacts of policies. Gender 'blindness' in policies has real world impacts; for example, respondents in Kenya highlighted how women in rural areas often have low influence over health decisions, so health policies need to consider the effects of this on women and children's access to health services. Respondents also highlighted how some Kenyan health policymakers tend to discourage gender-disaggregated reports as too 'bulky', leading to poorly targeted, and therefore ineffective policies. Inequalities also affect how stakeholder consultation is carried out, who is consulted and whose interests are represented in policies and their implementation.

The lack of a strategic focus on gender in the BCURE programmes meant that it was not possible for the evaluation to systematically explore gendered differences in programme outcomes to learn more about how gender and inclusion issues might affect the development of individual skills and organisational systems for evidence use in the longer term. In most of the monitoring documents reviewed for the evaluation, outcome data was not systematically gender disaggregated (e.g. to assess whether there were any qualitative or quantitative differences between the outcomes noted by male and female trainees in follow-up assessments). The evaluation sought to investigate whether differential outcomes had been observed for male and female participants, but no particular patterns emerged. Overall, the evaluation suggests that a gender perspective needs to be fully integrated into programme strategies and monitoring and evaluation from the start, if these dynamics and the implications for programme outcomes are to be more fully understood.

4. Three 'ways of working' that underpin success

The BCURE projects varied considerably, employing different combinations of activities and targeting different types of government stakeholders at varying levels of seniority. However, the evaluation found three cross-cutting 'ways of working' associated with greater impact across all six countries included in the evaluation, which underpin the three impact pathways discussed later in this report.

4.1 Going beyond needs assessments to think and work politically

BCURE was not initiated as a governance programme, and was initially framed by DFID as a technical rather than political intervention. Although all partners conducted needs assessments as part of their design phase, DFID did not require them to design interventions based on an assessment of the political economy of the target settings. However, the evaluation has highlighted the essential importance of understanding and responding effectively to political economy opportunities and constraints when attempting to promote evidence use. This chimes with recent insights from the broader EIPM literature, which emphasises the messy, political nature of evidence use in policymaking, and the importance of moving away from 'rational' understandings of policy processes towards a deeper understanding of the political and power dynamics that affect the extent to which evidence is used (see Box 3). It also resonates with the findings of the 2017 World Development Report, which emphasises that promoting better policies requires thinking beyond capacity building to consider power asymmetries that influence what is politically possible. While most BCURE projects took a technical approach at first, they began to be more alert to these dynamics as they progressed. Projects had greater success where they followed these principles:

- **Locating an entry point in a sector or government institution where there was existing interest in evidence, clear incentives for reform, and a mandate for promoting evidence use.** The most successful programmes formed partnerships with government institutions or units with a clear mandate to promote evidence use (Section 5) or a cross-government mandate for reform (Section 6). Success followed where there was a genuine interest in partnership, and where the incentives of the institution or unit aligned with BCURE's goals. The possibility of catalysing widespread change depends on how much power and influence the government partner has, as well as broader drivers and opportunities for EIPM at sectoral, ministry and national levels. The cross-government impact pathway describes how BCURE succeeded in rolling out government-wide reforms through its partnerships with 'heavyweight' government agencies. However, bigger is not always better. The parliament and single ministry pathways describe how BCURE had considerable success in small units, but struggled to make a dent across large, dispersed ministries. It is also important to consider what existing support is in place: what does the programme bring that's new? BCURE made a more limited contribution to capacity in contexts where there were a plethora of existing, much larger-scale capacity support programmes.
- **Taking an advantage of a window of opportunity – timing and prior relationships were often key**. All six projects had most success where they capitalised on national and local opportunities for partnership and reform. These windows of opportunity may be large scale (the adoption of a new constitution, an ongoing process of devolution) or more localised (a new senior leader with an interest in promoting better evidence use, a recently established unit in need of capacity building or looking for a flagship programme). Several BCURE partners built on existing institutional credibility and relationships to gain a foot in the door and access high-level champions, suggesting the value of partnering with credible, well-connected local institutions with an established national presence. Personal relationships should also not be underestimated – Section 5 highlights how connections were often made and buy-in bolstered through the networks, prior experience and credibility of individual staff members and consultants.

Box 3: Insights from the literature: Thinking and working politically

The literature review highlighted that **evidence use is infused with politics and power**, and that **evidence is just one part of a patchwork of factors influencing policy decisions**, alongside political and strategic considerations, expert opinion, stakeholder and public pressure, and resource constraints. Some writers have flagged that the literature on EIPM tends to be narrowly focused on 'barriers and facilitators' of evidence use, which encourages a technocratic diagnostic and discourages a focus on understanding how evidence is actually used within policy processes through active engagement with political and policy theories (Liverani, et al., 2013; Oliver et al., 2014).

In a recent book, Cairney (2016) picks up on these themes to argue that **EIPM is often based on an idea of 'comprehensive rationality'** that has been soundly debunked in contemporary policy theories – the idea that policymakers are able to make rational choices based on a clear set of preferences and an understanding of all relevant evidence. In the real world, policymakers make choices in a complex system, in which the role of evidence is often unclear, evidence (where it exists) does not hold easy answers about what to do, and the demand for evidence does not match the supply. Cairney suggests that actors seeking to promote EIPM need to consider and engage with actors, institutions, policy networks and broadly held ideas and norms, as well as the broader context within which decisions are made. This includes understanding policymaking 'sub-systems', learning what the 'rules of game' are within institutions and sub-systems, forming coalitions with other influential actors, and engaging in the policy process long enough to exploit windows of opportunity.

These insights from the EIPM literature point towards the importance of **considering ideas from political theory and the governance literature when designing programmes to promote evidence use**. The evaluation, like the BCURE programme itself, was not designed using a governance lens, and it has been beyond the scope of the evaluation to investigate this literature in great depth. However, some key recent ideas include:

- **'Thinking and working politically'** and '**doing development differently'**: recent literature emphasises the importance of international development programmes working 'with the grain,' being problem driven and locally led, understanding the political context through ongoing intelligence gathering, exploring the politics underpinning problems and opportunities for action, and adapting interventions as they go, by learning through experimentation (Andrews, et al., 2015, 2016; Booth and Unsworth, 2014; Faustino and Booth, 2014).
- **'Isomorphic mimicry**' is the idea that low and middle-income countries may commit to donor-sponsored public sector reforms in order to signal a willingness to modernise and access donor resources, but the resulting reforms are simply a hollow imitation of high-income country models, ultimately failing to change government practice (Matt Andrews, 2013). Other authors have unpacked this idea further – for example, Krause (2013) suggests that this kind of insincere mimicry is a rational response to the demands of donors and ultimately therefore a sign of a flawed development model. He suggests that there is a more serious danger of 'institutional ventriloquism' in fragmented, informal, often fragile states, where "*best practice reforms are articulated, planned and implemented following external prompting and via externally funded advisers and consultant*"' in the absence of "*whole of government intention*." In a recent paper by the Effective States and Inclusive Development research centre, Yanguas (2017) suggests that public sector reform programmes take root in a genuine way where there is a match between the ideas being proposed, and the norms and incentives of the ruling elite, and where there is space for reforms to become embedded given how competitive or dominant the political settlement is at a particular point in time. However, *"where there is a mismatch between types of policy ideas being proposed and ruling coalition incentives, the result may simply be the cosmetic adoption of best practices as a way of appeasing the international community or preventing reputational costs."*

- **Identifying individual champions within partner institutions and nurturing these relationships.** In several cases across all three impact pathways, senior officials acted as 'gatekeepers' to the institutional partnership. This was especially crucial in the cross-government pathway, where it seems unlikely that BCURE would have gained a foothold in influential national institutions without the enthusiasm and commitment of specific individuals. These champions also frequently acted as 'cheerleaders' for the programme, helping bring other senior stakeholders on board and identify further opportunities for partnership. They had a few things in common: they had significant seniority and power and they were personally committed to (often described as passionate about) reform. However, there are risks in tying a programme too closely to an individual champion, given the likelihood of senior staff turnover (discussed further in Section 6).

4.2 Thinking beyond 'skills' to build capacity at multiple levels of the system

All of BCURE's larger-scale success stories, as well as the non-BCURE impact case study, worked because they went beyond developing technical skills at an individual level. Three key insights are as follows:

- **Individual capacity is the bedrock for EIPM – but programmes need to think beyond building the knowledge and skills of policy and research staff.** BCURE was premised on the assumption that there are major capacity gaps in low and middle-income countries, around how to access, appraise and apply evidence effectively in policymaking. Partner needs assessments confirmed this, and individual capacity building has been a necessary foundational activity across all three impact pathways, mainly targeting the technical policy and research staff who are responsible for designing policy documents and developing research products that feed into policy formulation. However, one-off training is unlikely to be enough to embed new skills, as demonstrated in Spotlight 2 below. The single ministry pathway emphasises the importance of follow-up support to help trainees change their practices in the longer term. All three pathways also highlight the necessity of working with senior managers and government stakeholders. These individuals are generally not involved in the technical side of evidence access, appraisal and use, but their awareness and buy-in is essential to create an environment where technical staff are supported to work in a different way.
- **System-wide change requires a focus on the 'top down' as well as the 'bottom up'.** The BCURE programmes that have made the most progress towards catalysing longer-term change in evidence use have all employed 'top down' organisational-level activities as well as 'bottom up' capacity building. The cross-government pathway discusses how BCURE-supported EIPM tools and systems have helped catalyse change in line ministries, when backed by an institutional champion. The same approach has also worked within single ministry and parliament settings, when BCURE has supported the development of sector-specific tools or guidelines that have helped ministry or parliament officials use evidence more effectively. The single ministry pathway explores how accompanying government units through specific policy processes in an evidence-informed way can help to showcase the value of an evidence-informed approach.
- **It is important to ensure that interventions at different levels join up to have a system-wide effect.** While all of the BCURE programmes worked at multiple levels of the system through different activities, there is significant variation in the extent to which these activities meaningfully joined up with one another. The impact pathways discuss how some programme activities were too small scale, ad hoc and potentially too short to catalyse change within large ministries or parliaments, or too disconnected to combine to promote change at a national level. In some cases, programme resources were dispersed across a number of one-off and small-scale interventions, with limited cross-over in terms of the individuals involved. Where BCURE partners succeeded in joining up their interventions, results suggest greater potential of a longer-term catalytic effect – this also resonates with the literature on EIPM as a complex system, discussed in

Box 4. This suggests that programmes should also consider how interventions at different levels of a system can be meaningfully linked and coordinated with one another given the time and resources available, as otherwise there is a risk they will remain isolated rather than combining to create pressure for change.

Box 4. Insights from the literature: EIPM as a complex system requiring capacity at different levels

The BCURE literature review discussed a number of studies that frame EIPM as a complex system: a group of interacting, interrelated, and interdependent sub-systems and components that form a complex and unified whole (Coffman, 2007). Complexity theories hold that that change in 'complex' settings does not happen in a rational, linear way that can be predicted in advance. Instead, individual behaviours and interactions between people combine and amplify one another in diverse and sometimes surprising ways, with consequences that no one could have predicted (Smith and Joyce 2012; Ramalingam 2013).

If EIPM is a complex system, this implies the need to carefully consider interrelationships between individual, organisational and institutional capacity for EIPM rather than focus on one level at a time. For example, this might mean thinking about the influence of organisational systems on individual values, or the ways in which ideas about evidence in wider society shape how it is talked about, and which types of knowledge are considered important. This literature review highlighted the value of examining EIPM as a system, and considering capacity building for EIPM as a multidimensional issue. Multiple initiatives are likely to be needed, working together holistically over time, to support and catalyse capacity development (FAO, 2010; Capacity.org n.d.).

4.3 Accompanying change, not imposing it

BCURE programmes were more successful when partners 'accompanied' government partners through a flexible, tailored, collaborative approach that promoted ownership. 'Accompaniment' is understood within the evaluation as a general way of working that contrasts with a more traditional 'supplier/consumer' model, where an external partner provides ad hoc capacity support through one-off interventions. The value of this approach has been increasingly promoted in recent literature on adaptive programming and development entrepreneurship (see Box 5). Government reform processes are unpredictable and highly context-specific, meaning that it is not clear at the outset what will work. Accompaniment allows a more flexible approach, where government and external partners explore opportunities together in order to design approaches that have the potential to catalyse change.

Accompaniment has been adopted to different degrees and in different ways by various BCURE partners, but the evaluation consistently found that all of the BCURE projects were more successful where they adopted elements of this approach, and it was also an important reason for success in the non-BCURE impact case study. On a smaller scale, several partners accompanied ministries or units through specific policy processes, providing tailored and flexible support to co-produce new tools or policy products in a way that promoted government ownership (discussed in the single ministry and parliament impact pathways). On a larger scale, some projects adopted elements of the approach at a whole ministry or at cabinet level, working to promote reform through a collaborative model characterised by high levels of government ownership, spending significant time on building and maintaining senior-level relationships and buy-in, and responding to windows of opportunity where they arose. It is important to note that the ability of BCURE partners to work in this way has

been constrained by DFID's milestone-based contracting model, which has made it difficult for programmes to adapt their plans to changing contexts and has created challenges given the unpredictable and lengthy timeframes required when working with government. Partners have had to cope with regular changes in government personnel that require the rebuilding of relationships, as well as changes in political priorities that can block or accelerate demand for programme activities – all of which emphasise the need for flexible management arrangements.

Box 5. Insights from the literature: Accompaniment

There is emerging literature on the 'accompaniment' mechanism, from both the health policy and governance fields. The concept links closely with insights on 'thinking and working politically', detailed above, and also on 'adaptive programming' – the idea of moving away from top down objective setting and pre-set targets and towards adjusting programme activities and outcomes over time, through making 'small bets', getting fast feedback and scaling up successful interventions while halting unsuccessful ones (Andrews et al., 2016).

The 'Development Entrepreneur' approach developed by The Asia Foundation (Faustino and Booth, 2014) emphasises an approach to governance reform characterised by entrepreneurship, iterative 'learning-by-doing'; 'a close-knit team of passionate and determined leaders' united by 'a high-level of trust, a shared agenda and complementary technical and political skills', and finally programme staff having the time, space and authority to try out new and innovative ideas within established organisations. Faustino and Booth (2014), Booth (2014) and Cummings, (2015) discuss how this approach has played out in practice through various reform interventions, demonstrating that when an outside government agency is given the scope to provide flexible, collaborative and tailored support, this can promote 'learning-by-doing' at the local level, and facilitate strong partnerships that create the conditions for local partners to push for and own reform initiatives. Trust in and credibility of the external partner is viewed as essential to enabling this kind of partnership to flourish.

The literature on adaptive programming also emphasises the importance of flexible models of engagement, enabled by donor structures and reporting requirements that support flexibility and changes in direction in response to learning and new opportunities, rather than rigidly holding programmes to pre-set objectives and top down targets (Andrews et al., 2016; O'Neil, 2016). For example, Andrews et al. (2016) identified that successful health sector reforms had 'flexibility' where unsuccessful ones did not: they incorporated ongoing feedback on progress and results, created opportunities to adjust project content and incorporated new ideas into project activities based on ongoing learning.

Spotlight 2. How and why can training support evidence-informed policymaking?

All six BCURE projects used training or workshops to build individuals' knowledge, awareness and skills for EIPM. Overall, across most settings, the evaluation found strong evidence that these activities had increased participants' awareness, technical knowledge and/or skills. However, there were marked differences in the extent to which trainees had been able to apply their learning in order to access, appraise or use evidence more effectively. The findings are unpacked in more detail in relation to the three impact pathways, with further detail in Annex 9.

Where training led to behaviour change, this was a result of several mechanisms (or change processes) that often worked together. These mechanisms are drawn from widely held theories and concepts from adult learning and EIPM literature, discussed below.

- In Kenya, Zimbabwe and Bangladesh, where training was directly relevant to participants' day-to-day jobs, training (and follow-up interventions) built **self-efficacy** through imparting new knowledge and skills, and raising trainees' confidence in their ability to undertake a task or perform their roles.
- In Kenya, Pakistan, Sierra Leone, and Bangladesh, training (and associated technical support) provided participants with practical evidence tools or processes that **facilitated** them to do their jobs more easily or more efficiently– for example a policy brief format in Kenya, a set of EIPM guidelines in Bangladesh, and an introduction to cost-benefit analysis in Pakistan.
- In Kenya, behaviour change was also **reinforced** through demand for evidence application and use by managers and senior staff, which motivated participants to apply their learning. Recognition and rewards following from improved quality evidence products also generated further motivation for evidence use, stimulating a virtuous cycle. In Pakistan, where there weren't any such incentives to apply learning, the training appeared to have much less impact on behaviour change. The EIPM guidelines produced by BCURE in Bangladesh are also intended to reinforce behaviour change among trainees, through managers requiring them to follow the guidelines – however, it is too early to tell whether this will happen in practice.
- In Pakistan and South Africa, the training helped some participants **frame their thinking** on how to use evidence or data in order to address a specific task, providing them with new concepts or reminding them about the importance of evidence, in a context where they already had the technical skills and motivation to undertake the task. However, this mechanism was not associated with widespread behaviour change in BCURE projects.

In many cases training did not lead to demonstrable changes in evidence use, as a result of a broader environment unconducive to evidence-informed ways of working, and/or issues with training design and implementation (including reaching people with opportunities to use new skills). The evaluation emphasises the importance of beginning with a clear understanding of the political context, and the incentives and disincentives for evidence use within a specific setting – as described in Section 4. The findings are also consistent with widely held theories of adult learning, which suggest that training should be based on an understanding of needs, actively engage participants and build on their previous experience and motivation, and go beyond one-off training sessions to engage participants in a longer-term way (see below). BCURE training was more successful where:

- **Training aligned with incentives in the broader organisational and institutional environment.** Across the BCURE portfolio, learning led to behaviour change where there were incentives for participants to access, appraise and apply evidence. The attitudes of managers and senior staff was very important – the extent to which they encouraged and supported trainees to apply their learning, which in turn is related to broader incentives for or against evidence use within the civil service, when weighed up against partisan policymaking or widespread corruption.

- **Training was closely targeted to those who could apply it, and was directly relevant to their day-to-day work.** In Kenya and in the Ministry of Youth in Zimbabwe, relevance and applicability of training was ensured by targeting policy analysts and research officers who were required to search, access and appraise evidence as part of their roles. Trainees who had not put their learning into practice frequently said they had not had the opportunity to do so in their job – particularly in Pakistan, where the EIPM course is undertaken by all civil servants as part of broader mandatory training, rather than being targeted to specific cohorts and tailored to the type of work they do. Ensuring the 'right' participants take part in training is not always easy. It requires a detailed understanding of roles and incentives in units and divisions, not just at a ministry level; for example, in Kenya there were marked differences between units linked to differential opportunities and incentives to work in an evidence-informed way. In Bangladesh, BCURE developed rigorous criteria to ensure the training was only offered to civil servants directly involved in policy development, but in practice this had not always worked – managers are likely to have their own incentives for selecting certain participants, which need to be considered. The evaluation also highlights the importance of existing motivation to apply evidence, along with previous knowledge and experience in the concepts or tools introduced. These factors are often difficult to screen for.
- **Partners considered seniority and the trade-offs involved in targeting more junior or more senior staff.** Technical (usually mid-level or relatively junior) staff are generally responsible for creating policy and research products that feed into decision making, but may have limited autonomy to influence organisational shifts in evidence use. In both Pakistan and Zimbabwe, stakeholders felt that targeting mid-level and junior staff was more likely to make an impact than training senior staff, who are often political appointments and may be more resistant to change. However, senior civil servants often make the decisions, drawing on (or not) the evidence collated by their juniors, and may have most of the power – particularly in hierarchical civil service cultures – but are less likely to spend significant time accessing, appraising or applying evidence as part of their day-to-day work. Although the evaluation highlights the importance of support from managers in enabling trainees to apply their learning, this is not necessarily best achieved by mixing junior and senior staff, who will need different information and skills. It is also difficult to gain the commitment of senior staff to take part in multi-day training courses. Specific interventions for senior staff – such as a tailored half or full day course focusing on sensitisation rather than imparting technical skills – may be more appropriate.
- **Activities went beyond one-off training.** The projects with the most success at catalysing behaviour change went beyond one-off training courses to provide follow-up on-the-job support. For example in Kenya trainees were helped to develop policy briefs, and in Bangladesh ministry staff were assisted in policy development using new EIPM tools.
- **Training was practical and participatory, using local case studies or live policy examples.** Many trainees across all BCURE settings emphasised the importance of training being practical and participatory – using hands-on exercises to give participants an opportunity to practise skills. This resonates with the findings of the Stage 2 evaluation, which found that behaviour change was generally associated with increases in *skills* as well as knowledge. Making training participatory is not always easy in contexts where training is typically fairly didactic, and may require significant capacity support to local trainers. Trainees in Bangladesh, Kenya and Zimbabwe emphasised the importance of using local, sector or context-specific case studies, or giving participants the opportunity to work on live policy topics, to help relate examples directly to their work.
- **Training incorporated a focus on soft skills as well as technical skills.** In Kenya and Zimbabwe, BCURE training incorporated soft skills, which were frequently cited as some of the most important elements. EIPM is not a purely technical endeavour – it requires thinking about how to communicate research evidence, use evidence to influence decisions, and understand the needs of evidence users. In Pakistan, participants who had been able to use their new skills often said they had to present evidence and analysis to senior members of the civil service, requiring presentation skills and the ability to negotiate for resources.

- **Partners considered the experience and background of facilitators.** Many trainees emphasised the importance of ensuring facilitators were knowledgeable, patient and confident. This usually involved using local trainers rather than or alongside international trainers (potentially creating challenges in making training participatory, as discussed above). This may require facilitators who understand the specific sector as well as the broader national context – for example participants in the Parliament of Zimbabwe felt the facilitators were too 'academic' and did not understand the nuances of working in parliamentary settings.

Insights from the literature: Building individual-level capacity

Capacity development is about more than 'skills'. Several leading behaviour change researchers argue that behaviour change requires a combination of positive intention, skills and absence of environmental constraints (Fishbein and Middlestadt, 1994). Bryan et al. (2009) and Lyon et al. (2011) draw on the theories of andragogy and self-directed learning to suggest several 'key principles' of adult learning, including that adults need to know why they are learning and are motivated to learn by the need to solve problems; their previous experience must be respected and built upon; adults need learning approaches that match their background and diversity; adults need to be actively involved in the learning process; and they require extended contact (rather than one-off training sessions) in order to assimilate learning.

The evaluation has drawn on three key mechanisms from the literature on adult learning and capacity development:

Self-efficacy: One well-established learning theory, holding that individuals are more likely to behave in a particular way if they possess high 'self-efficacy' – in other words, if they believe they are capable of effectively performing a particular task or handling a particular situation. (Bandura, 1977). Empirical evidence suggests a link between self-efficacy, motivation and outcomes such as work attendance, productivity and future employment (Bandura, 1988; Eden and Avirma, 1993).

Facilitation: Walter et al. (2005) highlight facilitation as one of the key mechanisms underpinning interventions trying to promote research use. This involves enabling or facilitating staff to adopt EIPM behaviours, which led to the improvements in individual capacity and use of evidence in day-to-day work. This mechanism is underpinned by change management theories, which *"emphasise the importance of enabling strategies providing practical assistance for individuals and groups to change"* – for example technical, financial, organisational or emotional support.

Reinforcement: This is another key mechanism identified by Walter et al. (2005), involving "*using rewards and other forms of control to reinforce appropriate behaviour*". Reinforcement can be positive (e.g. some kind of reward or encouragement for evidence use), or negative (reminders, audit, mandatory requirements, performance assessments). Reinforcement is also a behaviourist theory of learning – suggesting that a learner will repeat a desired behaviour if positive reinforcement is given in the form of material or non-material rewards (Dunn, 2002).

Box 6. Cost-effectiveness of EIPM training

The structure, duration and scale of EIPM training varied considerably across the BCURE programmes. Some projects sought to roll out training courses across the whole of government (Bangladesh and Pakistan), others focused on specific ministries or units within a ministry (Kenya, Zimbabwe). Some delivered blended online and in-person training (Pakistan), others combined one-off training with ongoing coaching (Kenya). Because of these differences, the costs of the training, and the associated cost-effectiveness varied considerably.

Considerations of cost-effectiveness need to ask: did training improve knowledge and skills? Most importantly, did it lead to behaviour change? Looking at costs per participant in isolation can provide a distorted picture. Other factors which can drive cost-effectiveness of training include whether there is has been any cascading of training to others and whether the training has been institutionalised in a local training provider such as civil service training college.

	Bangladesh	Kenya	Zimbabwe	Pakistan
Total investment in training	£595,381	£204,975	£96,200	£1,056,960
Number of people trained	400	45	49	1,780
Cost per participant	£1,148	£3,950	£2,000	£594, reducing to £168 in the final year.
Evidence of change	Changes in knowledge and skills, but too early to judge changes in behaviour across the wider cohort of trainees. Some had applied learning through BCURE-supported pilot policies	Changes in knowledge, skills and behaviour	Changes in knowledge, skills and behaviour	Changes in knowledge, skills and attitudes, but concerns that majority of trainees will not have opportunities to apply new knowledge/skills
Reflections on cost-effectiveness	The cost-effectiveness of the training in Bangladesh depends on how far trainees are incentivised to apply their learning by the roll-out of the EIPM guidelines and if they are enforced by line ministries. If this happens, that could make the training cost effective given the reach in terms of numbers, otherwise the cost-effectiveness will be lower	There was good potential for the training in Kenya to be cost effective, as training contributed to skills improvement in both Parliament and MoH; however, in MoH there was less evidence of behaviour change, despite there being higher spending	There was good potential for the training in Zimbabwe to be cost effective, as training contributed to sustained improvement in skills and changes in behaviour in the MoYIEE; however, BCURE made a much lower impact in Parliament	Costs include the development of the platform as well as the delivery, so efficiencies in the cost per participant were achieved as greater numbers were reached. There was potential for the training in Pakistan to be cost effective given the reach in terms of numbers and the evidence that training contributed to learning gains; however, this potential was reduced as the relevance to the majority of participants was low, which prevented skills application and behaviour change

5. Impact pathway 1: Support to a single ministry

BCURE worked to promote EIPM within specific ministries and departments in Bangladesh, Zimbabwe, Kenya, Pakistan and South Africa. This section also draws on the non-BCURE impact case study of the South Africa National Evaluation System. In Bangladesh, BCURE's work within specific ministries formed part of its broader cross-government approach – the findings are therefore discussed both in this section and the next. BCURE's work in single ministry settings was fairly diverse, ranging from tailored support to research units, to assistance in developing pilot policies and data decision tools – described in detail in Box 7 below.

Box 7. Summary of BCURE single ministry support

Two approaches: 'Training-plus' and 'support to policy pilots'

In **Zimbabwe**, the evaluation focused on BCURE's work with the Ministry of Youth, Indigenisation and Economic Empowerment (MoYIEE). BCURE adopted a 'training-plus' approach with the newly established Research and Policy Coordination Unit, delivering a training course followed up by various organisational-support activities including policy dialogues, support to a terms of reference for the unit, and a series of organisational change workshops.

In **Kenya**, BCURE adopted a 'training-plus' and 'support to policy pilots' approach with **the Ministry of Health** (MoH). The MoH had established a small internal Research and Development (R&D) unit in 2013 to promote access to and use of research within the ministry, which provided an important entry point for BCURE. The programme provided a package of support to technical officers and mid-level managers across the MoH, consisting of an EIPM training course, follow-up coaching, co-production of a set of EIPM guidelines, and support to a Research for Health (R4H) policy.

In **Bangladesh**, BCURE worked with three ministries, and the evaluation focused on the work with the Ministry of Commerce (MoC) and Ministry of Environment and Forests (MoEF). The approach was also 'training-plus' combined with 'support to policy pilots', but slightly different to Zimbabwe or Kenya as support to ministries was nested within an overarching cross-government approach (discussed in more detail in Section 7). The programme trained around 100 officials responsible for policy formulation in each ministry, and facilitated working groups to develop 'pilot policies' using the EIPM guidelines developed with Cabinet Division.

In **Pakistan**, BCURE was unique in working with provincial level government departments, adopting a 'support to policy pilots' approach. The evaluation focused on its work with the Punjab Excise and Taxation (E&T) Department, Sargodha District Police, and the Punjab Health Ministry. BCURE led a series of policy pilots to develop digital data-driven decision-support tools that could help front line workers (tax-collectors, police and community health workers) improve service delivery.

In **South Africa**, BCURE adopted a 'support to policy pilots' approach with the Department of Planning, Monitoring and Evaluation (DPME). While the project undertook a number of activities with various Ministries (see Section 3), this report focusses on its support to co-produce a tailored evidence mapping product designed to assist policymaking.

Non-BCURE case study of the South Africa National Evaluation Strategy

As described in Section 2.2, the evaluation incorporated a non-BCURE case study of the South African National Evaluation System (NES), which was established in 2012 to promote and support evaluations across government. The case study examined two of its completed evaluations: a Diagnostic Review of the Early Childhood Development policy, and an evaluation of the Business Process Services programme. DPME is the custodian of the NES, and provides various support to departments undertaking evaluations – including co-funding, developing standards and guidelines, delivering training on evaluation, providing hands-on support and advice through active participation in evaluation steering committees, and ensuring that findings are addressed through Implementation Plans.

5.1 What were the drivers, opportunities and risks for EIPM in ministry settings?

The BCURE single ministry contexts had several challenges and opportunities in common:

Political environments were to varying degrees restrictive and fragmented, with concerns around mismanagement and corruption. Across the BCURE operating countries, the political environment places varying degrees of restriction on freedom of expression, public voice and civil society. Contested and fragmented political landscapes are reflected in highly politicised ministerial and sectoral contexts, where policies are often perceived to be partisan rather than evidence based, restricting space for individuals within government to challenge the status quo. Many of the settings are also characterised by high levels of corruption, leading to resource allocations being directed towards political allies and constituencies.

Capacity shortcomings, financial constraints and staff turnover also create difficulties with embedding EIPM at a unit or departmental level. All the settings face ongoing capacity shortcomings in the civil service to develop and implement policy. This includes persistent lack of access to data by various stakeholders, including government personnel, and poor quality data collected by devolved or provincial public services for monitoring. There are also financial constraints and reliance on donors to support programming – inhibiting ability to increase budgets for research and data collection. In several contexts, rapid change in senior leadership, and high levels of staff rotation, limit opportunities for EIPM to take hold, with new officials unfamiliar with previous initiatives or promoting their own agendas.

Alongside these challenging conditions, there were a range of high-level initiatives emerging in the BCURE sectors and ministries that created opportunities for enhancing evidence use. These included the strong role of international donors in government programming, which created a drive for evidence and robust monitoring and evaluation (M&E) to support improvements in performance and efficiency of partner governments. In response to political, citizen and donor pressures for improved effectiveness, governments have established high-level initiatives around results and evidence, including results-based management systems, public service charters and International Standards Organisation (ISO) quality standards; and strengthened national and sectoral planning and M&E processes, albeit moving at different paces. Linked to this, there is growing momentum for the use of technology and digitisation to improve government information systems, especially in Pakistan and Bangladesh where there are ongoing initiatives and existing government bodies working on these issues. The increased transparency that comes with technology is perceived as feeding a growing public interest in policymaking, emphasised by public protests or sector strikes, which in contexts like Zimbabwe is linked to a sense of an urgent need for the government to find solutions to national challenges.

Finally, there are pockets of opportunity for EIPM in certain sectors where the political stakes are lower. This can be found for example, where there is a drive to improve performance because of a sector's economic and political importance – health in Kenya and commerce in Bangladesh – and supportive leadership, for example in the Bangladesh MoC. In some countries research and data are emerging as strategic investment areas, including in priority sectors such as health in Kenya, leading to the establishment of research and policy units within ministries with a mandate to promote EIPM and research use.

5.2 What happened and why?

Figure 4 depicts how BCURE worked towards a step change in evidence use through providing support to single government ministries or units. It does not represent what any individual project did, but rather synthesises evidence from across BCURE and the broader literature on how and why capacity support can lead to change. The diagram summarises what the evaluation has learned about how capacity support can contribute to EIPM in a single ministry setting, but it is not a fully tested theory, as projects made different degrees of progress towards the intended impact. It is therefore intended as a broad 'road map' for future programmes working to promote EIPM in single ministry settings, rather than a definitive prescription.

Below the diagram, Table 3 presents an overview of how far each of the outcomes were achieved by the BCURE projects (EQ 1) and BCURE's contribution (EQ 2). These findings are unpacked throughout the rest of this section, which also explores how and why BCURE did or did not make a difference (EQ 3).

Figure 4. Single ministry impact pathway

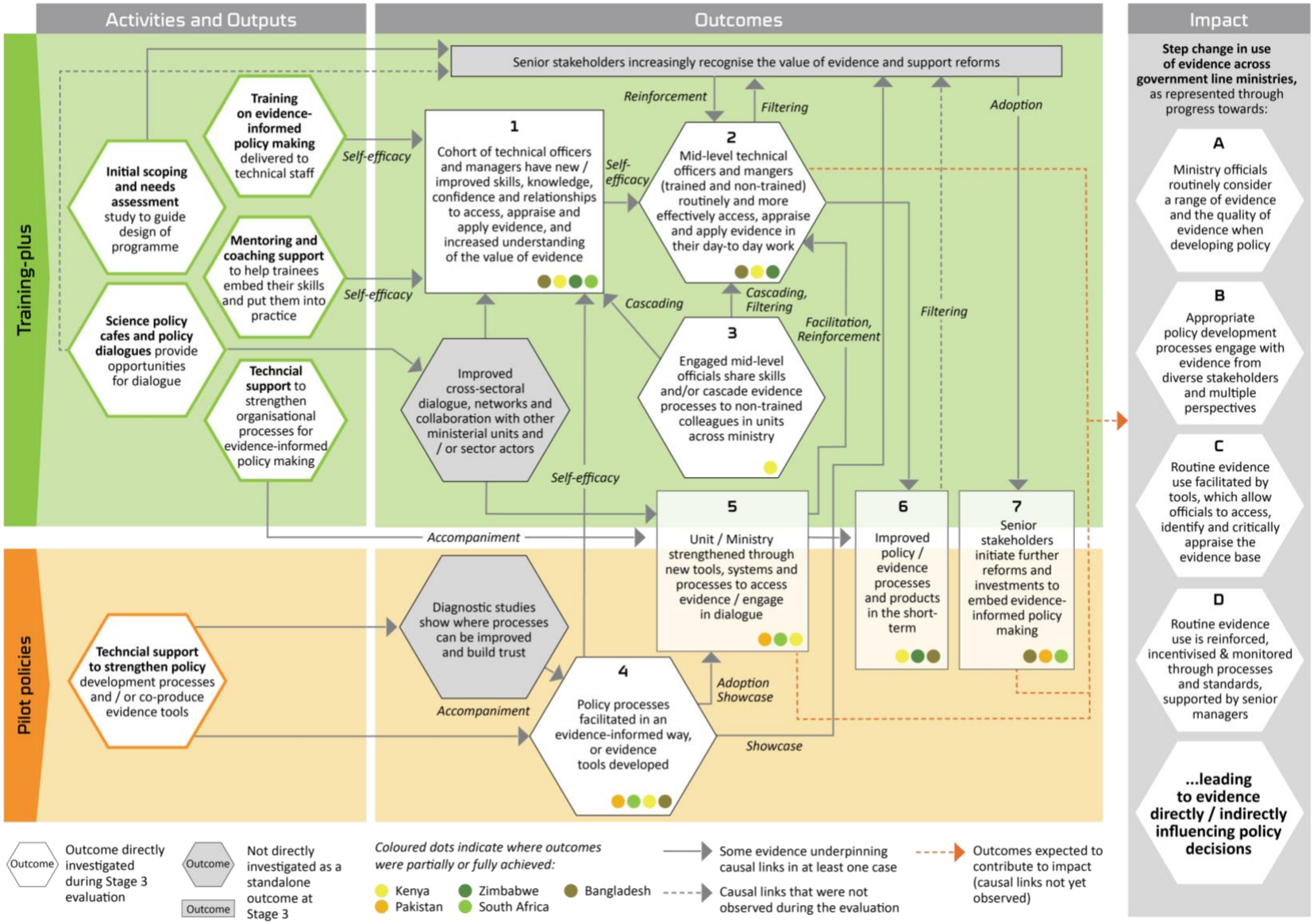


Table 3. Summary of evidence for single ministry impact pathway

Outcome	Summary of evidence for outcome (EQ 1) and BCURE contribution (EQ 2)
'Training-plus' approach: Bangladesh, Kenya and Zimbabwe	
1. Cohort of technical officers and managers have new/improved skills, knowledge, confidence and relationships to access, appraise and apply evidence, and increased understanding of the value of evidence	Strong evidence that trained staff gained new or improved knowledge, skills and confidence to use evidence in **Kenya, Zimbabwe** and **Bangladesh**, and that BCURE made a crucial or important contribution to this, in some cases alongside other donor-supported capacity building projects. Some evidence from **South Africa** and **Bangladesh** that the BCURE-supported evidence map and pilot policies also built the skills and confidence of ministry staff.
2. Mid-level technical officers and managers (trained and non-trained) routinely and more effectively access, appraise and apply evidence in their day-to day work	Strong evidence in **Zimbabwe** and **Kenya** that BCURE had made an important contribution to evidence use among many trained staff, but in both cases this fell short of routine use across key areas of their work and, in **Kenya**, was limited to trainees using evidence in specific units and divisions rather than across the ministry as a whole. Strong evidence in **Bangladesh** that staff supported by BCURE to apply EIPM guidelines in a pilot policy process had accessed, appraised and applied evidence more effectively as a direct result, but limited evidence that trainees not involved in pilot policies had applied their learning – although this is too early to definitively judge given the stage of the project, and it was only possible to interview a small proportion of the overall training cohort. No evidence from any setting that non-trained technical officers or managers were more routinely using evidence as a result of BCURE.
3. Engaged mid-level officials share skills and/ or cascade evidence processes to non-trained colleagues in units across ministry	This outcome was relatively implicit rather than a major part of project designs, but was a necessary step for projects to have a broader impact within ministries. As yet, there is no evidence that officials have shared or cascaded new skills or processes to non-trained colleagues or managers in **Zimbabwe**, or that officials in the **Kenyan** MoH have influenced the behaviour of colleagues or managers in the ministry. However, there is strong evidence of an unintended outcome from **Kenya**, in which mid-level officials cascaded BCURE EIPM training to non-trained colleagues at county level, as a direct result of the programme. The **Bangladesh** programme had a slightly different assumption around critical mass, discussed in the cross-government impact pathway.
'Policy pilots' approach: Bangladesh, Kenya, Pakistan and South Africa	
4. Policy processes facilitated in evidence-informed way, or EIPM tools developed	Strong evidence in **Kenya** that BCURE made an important contribution to supporting evidence use within the Research for Health (R4H) policy process using BCURE-supported EIPM guidelines, although the process is not yet complete. Some evidence in **Bangladesh** that BCURE made a crucial contribution to supporting pilot policies to apply evidence more systematically, again using EIPM guidelines – however, the case study only examined two of six supported processes, the process was less successful in one of the two ministries examined for the evaluation, and the policies have not yet been approved. Strong evidence in both **South Africa** and **Pakistan** that BCURE made an important contribution in supporting the development of good quality decision-making tools that drew on data and evidence.

Outcome	Summary of evidence for outcome (EQ 1) and BCURE contribution (EQ 2)
Longer-term change across both approaches	
5. Unit / ministry strengthened through new tools, systems and processes to access evidence / engage in dialogue	Strong evidence in **Kenya** that the programme had directly supported the creation of good quality EIPM guidelines that were officially adopted by the MoH as part of its draft standard operating procedures, but there is no evidence that these guidelines have been used beyond the BCURE-supported R4D policy. Limited evidence from **Zimbabwe** that the MoYIEE research unit has been significantly strengthened through new tools, systems and processes to access evidence. It is too early to judge this outcome for **Bangladesh** as the EIPM guidelines have not yet been adopted at ministry level, although the guidelines do offer the potential to strengthen policy development processes (see outcome 6). Strong evidence from **Pakistan** and some evidence from **South Africa** (from programme reports completed after the Stage 2 evaluation) that the BCURE-supported EIPM tools have been adopted within targeted government units.
6. Improved policy / evidence processes and products in the short term	Strong evidence in **Kenya** that BCURE support made an important contribution to improving the quality of the R4D policy. Some evidence in both **Kenya** and **Zimbabwe** that a number of trainees were using their skills to inform policies or generate evidence products, and that BCURE had made an important contribution to this, alongside including donor-supported programmes providing an opportunity for evidence use. However, in both cases the examples are mainly ad hoc. Strong evidence in **Bangladesh** that the MoC pilot policy used the EIPM guidelines effectively and resulted in a high quality product – although the MoEF policy was less successful, using the guidelines only partially. Both policies succeeded in being submitted to Cabinet although at the time of the final evaluation had not yet been approved. Limited evidence from **Pakistan** that the decision-support tools were as yet leading to more evidence-informed decision making. It was not possible to follow up the evidence map process in **South Africa** to investigate its effects on policy processes, as this country was not included in the Stage 3 evaluation.
7. Senior stakeholders initiate further reforms and investments to embed EIPM	Limited evidence from **Kenya** or **Zimbabwe** that BCURE influenced senior ministry stakeholders to initiate further reforms and investments to embed EIPM through the 'training-plus' approach. Some evidence from **South Africa** and **Pakistan** that policy pilots succeeded in catalysing further resources to scale up the new EIPM tools, although this is on a relatively small scale as yet. BCURE made some contribution to this alongside other factors – in particular, pre-existing high-level interest in data visualisation tools, and the existence of institutions that could leverage resources for scale-up. There is also some evidence that BCURE contributed to this outcome in **Bangladesh**, discussed in the cross-government pathway.
Impact: Step change in use of evidence across ministry or unit	Limited evidence from any setting that BCURE has promoted a step change in evidence use within target units or ministries. However, there is some indication of pockets of success across all the settings, where BCURE made an important or crucial contribution to improving the use of evidence within specific policy processes using EIPM tools (**Kenya** and **Bangladesh**), developing tools to facilitate evidence access and appraisal which were adopted by government (**South Africa** and **Pakistan**) and capacitating a new research unit that can continue to develop evidence products into the future (**Zimbabwe**). A number of ad hoc examples of change were also observed in **Zimbabwe** and (to a greater extent) in **Kenya**, where individuals were able to apply skills gained through BCURE training to improve evidence use within specific policy processes, but where this fell short of routine evidence use.

'Training-plus' approach

In Kenya, Zimbabwe and Bangladesh, BCURE provided capacity support to individuals alongside or followed up by technical support to promote organisational reforms. Although this did lead to improvements in knowledge and skills (**O1**), and some behaviour change among targeted trainees, as yet there is limited evidence that participants are using evidence in a sustained, routine way (**O2**). Organisational tools produced through technical support have the potential to strengthen EIPM within ministries, but have not yet been widely used (**O5**).

Table 4. Summary of 'training-plus' support

Country	'Training-plus' support provided
Bangladesh	Around 100 officials involved in policy formulation were trained in each of the MoC and MoEF. Training was structured around a set of EIPM guidelines, developed in conjunction with Cabinet Division (as part of the cross-government approach in Bangladesh, discussed in Section 7). In parallel to the training, the programme provided support to two 'pilot policies' in each ministry, coordinating working groups to develop policies using the EIPM guidelines, and providing national and international consultancy support.
Kenya	35 mid-level officials were trained in the MoH. This was followed up by individual and group mentoring to help trainees apply their new skills to develop policy briefs. The programme also provided technical support to the ministry to develop a set of EIPM guidelines. The EIPM training curriculum was revised in collaboration with MoH stakeholders and disseminated alongside the guidelines.
Zimbabwe	12 policy and research officers were trained in the MoYIEE, including all six members of the new research unit. This was followed up by ongoing technical support to help two trainees put 'Action Plans' into practice, which aimed to tackle organisational-level barriers to evidence use. The VakaYiko programme also produced an EIPM toolkit, containing training and other EIPM resources, which was distributed to the trainees.

There is strong evidence that BCURE participants in Kenya, Zimbabwe and Bangladesh all acquired new knowledge and skills that led to some degree of behaviour change, although as yet there is limited indication of *routine* evidence use among trainees. The three programmes predominantly targeted mid-level, technical officials, more likely to have a hands-on role in policy development. As discussed in Spotlight 1 above, trained officials were more likely to change their behaviour when their roles required them to engage with evidence, and when their team or unit provided the incentives, opportunities and resources for evidence use in policy processes:

- In **Kenya**, around half the training cohort suggested they were using evidence more routinely after the end of the programme, but a number highlighted limited or no real change in other areas of their work. This was because there had been limited opportunities for them to apply their skills, as policy processes typically take a few years and tend to be driven within donor programmes. The most significant examples of sustained evidence use were provided by trainees involved in developing policies and guidelines in key donor-supported sectors. These offer opportunities for trainees to apply evidence skills, as donor programmes tend to be data-driven and have a performance-oriented culture, as well as bringing into the MoH the technical and financial resources needed to undertake a policy or standards development process. In these contexts, the follow-on support to help trainees develop policy briefs seemed to help embed

skills because trainees were working on briefs that directly supported their policy work, e.g. a brief on water quality surveillance formed the basis of an environmental health policy proposal. Where trainees had applied their new skills in these contexts, this seemed to trigger positive reinforcement: by using evidence more effectively, motivated individuals improved their performance, which made them more likely to be promoted because good performance scores made them eligible for a wider range of roles. In addition, credible, evidence-based policies can help to inspire confidence in donor partners and facilitate resource mobilisation, itself a key performance target for MoH officials. All these positive incentives seemed to reinforce evidence use and helped it to become more routine for these individuals.

- In **Zimbabwe**, officials and managers within the MoYIEE's research unit felt that trainees were still applying the skills they learned through BCURE training two years later, and that the training had enhanced the quality of their work, boosting their confidence and soft skills including in communicating with their superiors. This was because the training had enabled trainees to fulfil their roles as research officers, in a resource-constrained context where other capacity support was unavailable, and in a setting where senior managers were interested in evidence and supportive of them applying their skills. However, the work of the research unit was constrained by resources and by its relatively junior position within the ministry, meaning that staff were frequently engaged in administrative rather than research tasks. As in the Kenya case, where trainees had been able to use their skills this was often through donor-supported projects.
- In **Bangladesh**, while there was evidence that the whole cohort of trainees had gained new knowledge and skills, the main evidence of behaviour change was among staff who had been involved in the policy pilots. Although it is too early to make a judgement about this given that many participants had only recently completed the training, in some cases staff did not think they would have an opportunity to apply the training in their roles, despite the fact that BCURE had a clear selection process to avoid this scenario. As in other countries, this was potentially a case of people being relevant 'on paper' but not in practice. The Kenya and Zimbabwe cases suggest that routine evidence use in future is likely to depend on how far BCURE succeeds in generating top down incentives for staff to work in an evidence-informed way, through its work with Cabinet Division – discussed further under the cross-government impact pathway.

The follow-up technical support provided through the 'training-plus' approach had mixed results. In Kenya and Bangladesh, good quality EIPM guidelines have been produced that offer the potential to strengthen units or ministries, but as yet there is limited proof that this has influenced evidence use beyond BCURE-supported policies. In both cases the guidelines were user-friendly, providing a structured and stepwise approach to the policy development process and how evidence should be searched for, appraised and applied.

In Kenya, guidelines were produced collaboratively by officials from the R&D unit and the Policy and Planning Division, with technical support from BCURE, a factor that was highlighted by respondents as critical in achieving institutional ownership by the MoH. The guidelines were completed and signed off by the Cabinet Secretary in 2016, in a context where the ministry was preparing itself for ISO quality certification, which required the MoH to have standard operating procedures. They were used successfully in the Research for Health policy process that BCURE accompanied, and also stimulated a decision to develop further guidelines – discussed further below. However, a year later, the final evaluation found no evidence that MoH officials were still using the BCURE-supported EIPM guidelines, or were even aware of their existence, and the policy process guidelines had not been finalised. The main reason was that the R&D unit drastically reduced in size at the end of 2016, leaving just one official with no budget or resources to champion the use of the guidelines internally. Despite recognition that guidelines could help improve the quality of work, they were never made mandatory, and without resources to promote them they were unlikely to be spontaneously adopted in a setting crowded with a plethora of existing policies, guidance and standards –

especially as they were never formally disseminated except through email. They were due to be launched at a high-profile health summit, but strikes by health workers and allegations of corruption at the ministry meant this event was postponed. Although the R&D unit has tried to mobilise resources to formally roll out the guidelines, efforts have been unsuccessful so far.

> *There was supposed to be dissemination [of guidelines], and training of trainers, and at county levels so that they can cascade it down…But training has not gone very far, there were very few counties that were trained. It is about difficulties with getting resources. (BCURE Participant, MoH, Kenya)*

The Kenya case perhaps provides a cautionary tale for Bangladesh, which is at an earlier stage in its work to establish EIPM guidelines. In the short term, the draft guidelines proved useful in the BCURE-supported policy pilots in Bangladesh, helping generate good quality, evidence-informed policies – discussed in more depth below and in the cross-government impact pathway. However, it remains to be seen how far the guidelines will be adopted in a meaningful way by ministries, and how far they will continue being used now that the programme has ended. In addition, while the Kenya and Bangladesh cases demonstrate that guidelines can be an important part of a broader capacity support programme (also seen in the South Africa impact case study, where DPME-produced evaluation tools and templates helped guide ministries through the process of conducting quality evaluations), the proof is in their use rather than their adoption. There is a risk that guidelines will be adopted on paper but not in practice – an example of 'isomorphic mimicry' (see Box 3 above).

Box 8. Summary of what worked for who and why in building organisational capacity through 'training-plus' (EQ 3)

In Kenya, Zimbabwe and Bangladesh, supporting individuals to use evidence more effectively through training and mentoring (I) succeeded in building self-efficacy (M), because training targeted cohorts of mid-level, technical officers who had some hands-on involvement in policy development (C). This led to improved knowledge and skills (O1) and more effective use of evidence (O2). In Zimbabwe and Kenya, where trained officials were able to use evidence in a more routine way (O2) this was because their roles required them to, and their units provided the opportunities, resources and incentives for policy development processes, mainly as a result of donor-funded programmes (C). In Kenya, evidence use led to better quality policy products (O6), for which officials were positively rewarded (C), reinforcing evidence use (M), and helping it to become more routine for these individuals (O2).

In Kenya and Bangladesh BCURE succeeded in developing EIPM guidelines through a collaborative process (I) of accompaniment (M) that created a quality stepwise product to support policy development (O5). In Kenya, this prompted the MoH to adopt the guidelines as official procedures, albeit voluntary not mandatory. However, they have not been used to support routine evidence use (O2), mainly due to a lack of resources for promotion and roll-out (C), in a context where there are many existing guidelines with no organisational platforms to enable officials to access them (C). It is too early to judge whether guidelines will be meaningfully adopted by ministries in Bangladesh (O5), although the Kenya case suggests this may not be straightforward. In Zimbabwe, although organisational support was provided to help embed the learning through training, this was through small-scale, ad hoc activities rather than accompanying the ministry in a flexible way (I) and did not strategically engage senior staff in order to leverage buy-in and resources (I) – as a result the project did not generate organisational tools that substantially strengthened the research unit (O5) or facilitated routine evidence use (O2).

In Zimbabwe, there was limited evidence that BCURE's follow-up technical support had strengthened the research unit. This was partly because the technical support was reduced in scope due to changing priorities and funding challenges, and partly because activities were designed as relatively ad hoc interventions that did not strategically build on one another, and were delivered by BCURE through a 'supplier/consumer' rather than 'accompaniment' model (discussed in Section 5). Crucially, although the technical support activities were designed based on trainee Action Plans, and hence in theory had ownership and buy-in within the unit, constraints in implementing partner capacity limited the ability of the project to provide strategic and joined up support. The programme also did not gain the ongoing engagement of senior ministry staff, and this appeared to limit the impact of reforms.

There were few signs that trained officials had shared their skills or cascaded new evidence processes to colleagues, other than one example from Kenya (**O3**). There was also limited evidence that changes in practice among technical staff had 'filtered up' to influence senior stakeholders, or sparked further investments in research or EIPM capacity (**O7**)

Although BCURE succeeded in promoting behaviour change among targeted technical staff in Kenya and Zimbabwe, there was limited evidence of this leading to change in attitudes or behaviours among trainees' colleagues or high-level stakeholders. In both contexts, the BCURE projects had an implicit rationale that the training-plus model would establish a 'critical mass' of individuals who were using evidence more effectively, which would in turn influence colleagues not involved in the programme. In Kenya, the assumption was that trainees would cascade skills within their teams, and also play a role in influencing managers of the benefits of evidence use. In both countries, the hope was that hosting the BCURE programmes in internal research units would strengthen their capacities to promote EIPM across the ministry, and that this in turn would attract senior stakeholders to invest more resources in the units to sustain the evidence agenda into the long term. However, there is limited evidence that this happened in either case. The 'critical mass' phenomenon is widely documented in social sciences (see Box 9), but several factors seem to have prevented it from operating in Kenya and Zimbabwe:

- **Failure to train sufficient numbers or to 'cluster' trainees effectively.** In the large MoH in Kenya, there were simply insufficient numbers trained, for too short a time, and they were too scattered throughout the organisation. In a context where divisions and units operate effectively in silos, this meant that trained individuals were too isolated to support each other or form clusters to promote new evidence-informed practices. Respondents in Kenya suggested that the programme should have had an explicit 'clustering' strategy, where a number of trainees could be clustered together in the same units to support each other and become 'focal people' for evidence use, then supported by BCURE to cascade skills to colleagues. This was broadly similar to the approach taken in Bangladesh in the pilot ministries (see the cross-government impact pathway for a detailed analysis).
- **Insufficient engagement of senior stakeholders.** In both cases, senior managers had not been engaged in training or sensitisation activities, and there was some evidence from Kenya that they may not have understood the new skills that trainees were bringing, making it unlikely that they would allocate time or resources for skills-sharing. In Kenya, a number of respondents felt the programme could have done more to engage senior managers directly; for example, through a half-day version of the EIPM course (an activity that was in the original programme strategy but was dropped due to resource constraints). Similarly, in Zimbabwe, BCURE's failure to engage senior staff was part of the reason the project did not generate broad support for organisational reforms outlined in trainees' Action Plans.
- **Rapid staff turnover and high-level incentives that dampen senior demand for EIPM.** Insufficient engagement of senior staff is only part of the picture. In both countries, wider

political economy factors are likely to have had a strong influence on senior managers' attitudes and behaviours towards evidence use, inhibiting the potential for changes in practice among technical staff to 'filter up'. In Zimbabwe, the directors in the Youth Development department (in which the research unit sits) were reportedly supportive of EIPM. However, the MoYIEE is widely viewed as one of the most politicised ministries, in an authoritarian context where it is safer not to challenge the status quo – this provides little incentive for senior ministry staff to engage with evidence that may not support the accepted political position. Rapid turnover in senior leadership has also resulted in new high-level stakeholders taking up positions of power who are unaware about EIPM and are therefore less likely to understand and be interested in the work of the research unit. For example, in 2014 a new minister came into post and introduced a parallel approach to developing a youth investment case, overriding the (evidence-informed) process the research unit were involved in. In Kenya, the evaluation found that evidence use can be incentivised if it helps improve performance in sub-sectors such as water and sanitation, where, if technical officers use evidence more skilfully and more routinely, this can enhance the professionalism and performance of the unit, inspiring confidence in donors to invest in health programmes. However, as in Zimbabwe, undesirable incentives come into play at senior levels. Given the scale of resources going into health, coupled with weak oversight and procurement systems, this can negatively incentivise evidence use as a means of gaining control of large-scale funds for personal gain, political influence and corrupt behaviours.

Box 9. Insights from the literature: unpacking the concept of critical mass

Several BCURE programmes were underpinned by the theory that reaching a 'critical mass' of officials in targeted government units would lead to changes in practice that would diffuse out to influence colleagues' behaviour. This concept has a long history in the social sciences, linked to models of collective action and diffusion of innovations (Greenhalgh et al., 2004; Rogers, 2003). The evaluation has attempted to unpack the different ways in which the BCURE projects were hoping to catalyse a critical mass effect, including:

- **Cascading out:** BCURE participants formally cascading their learning through introducing new ways of working or new structures and processes within their organisations.
- **Filtering out:** Changes in the way BCURE participants' access, appraise and apply evidence leading to recognition of the value of an evidence-informed approach among their colleagues, in turn influencing their behaviour.
- **Filtering up**: Improvements in evidence access, appraisal and use leading to higher-level recognition of the value of an evidence-informed approach, through senior staff seeing and being impressed by the benefits that EIPM can bring, and this in turn leading to increasing support (and resources) for EIPM at a senior level in the organisation.

The literature suggests that achieving 'critical mass' in social settings depends on a number of contextual factors, including the size of the cohort of adopters of a new practice within a setting, how connected they are with each other, organisational incentives or disincentives to give time, resources and senior management support to sharing new skills and approaches, and there being a clear advantage to changing behaviour such as quality improvement or career rewards (Greenhalgh et al., 2004; Rogers, 2003).

- **Limitations in the resources, power and positioning of research units.** Both Kenya and Zimbabwe had relatively new research units within the target ministries, which offered a good entry point and home for the proposed programmes. However, both units were constrained by reduced staff numbers and a lack of resources, which inhibited them from establishing wider influence within their ministries and sectors. Many respondents felt that cascading was not possible in this context without additional donor support (although this could have been a reflection of the 'donor dependency' culture as much as real constraints). In Zimbabwe, the research unit mainly works in one section of the ministry, with limited interaction with other areas, further limiting its potential to influence ministry-wide culture or approaches to evidence use. Senior officers make the decisions about what policy proposals are put forward for consideration, with the research unit playing a technical (and often administrative) supporting role rather than actively proposing policy alternatives.

Overall, the failure of the BCURE programmes to activate the 'critical mass' effect limited the sustainability and impact of the training in the Zimbabwean MoYIEE and Kenyan MoH.

There was one important example where EIPM training was 'cascaded' in the Kenyan MoH. This was not an intended outcome, but provides an example of how a cascading strategy could work given the right conditions. A group of mid-level managers in the MoH, once they had been through the first round of EIPM training, saw the opportunity to deliver EIPM training to county health officials. They adapted the curriculum, mobilised financial support from a separate funder, and successfully trained 65 people in four counties, more than the numbers trained through the original BCURE course. This happened as a result of the MoH's institutional mandate to build the capacity of county health administrations, along with significant pressures to deliver this in the context of a fraught devolution process. Together with the availability of funds for county-level training, this meant that it became easier to cascade the training outside the MoH than within it. The BCURE team was only able to provide limited support to this process, due to the constraints of their contract.

Box 10. Summary of what worked for whom and why in building a critical mass for change (EQ 3)

In Kenya, EIPM training was cascaded (M) to county officers (O3), through existing connections between the trainees and the county-level staff (C), in a context where there were clear advantages of cascading the training to the counties as it met the MoH's mandate (C), and (non-BCURE) resources available to support the new initiative (C). However, cascading did not happen in the MoYIEE in **Zimbabwe**, despite the fact that several participants felt it was important to share learning with provincial officers – this was inhibited by a lack of resources to conduct additional training in a context of severe national economic challenges (C), and an associated expectation by participants that such activities need to be funded by donors (C).

In both countries, there was an implicit assumption that improvements in the quality of work of junior staff would 'filter up' to influence senior staff (M), who would see the quality of the work and subsequently offer support to further EIPM capacity building and reforms (O7); or that changes in behaviour would 'filter out' to influence the behaviour of colleagues (O2). This did not happen in either country. In **Zimbabwe** this was in part a consequence of the relatively junior role of the research unit in the organisational hierarchy of the MoYIEE (C) and high turnover of senior-level stakeholders (C), while in **Kenya** trainees were too few and too scattered across a large department (I). In both cases, BCURE did not sufficiently engage senior staff (I), and, crucially, filtering up was inhibited by limited incentives for senior staff to challenge the status quo using evidence (C), or incentives to use evidence strategically to gain access to resources (C). The relatively small size and limited influence and resources of the research units (C) also inhibited the potential for a critical mass to emerge, particularly in the highly politicised environment of the Zimbabwean MoYIEE.

'Pilot policy' approach

In Kenya and Bangladesh, BCURE provided technical support to policy development processes to showcase an evidence-informed approach, while in Pakistan and South Africa, BCURE-supported specific tools designed to facilitate evidence use in decision making (**O4**). All of these activities supported evidence use in the short term (**O5**), and in most cases improved individuals' skills and confidence (**O1**), but had varying degrees of long-term impact.

There is strong evidence that BCURE succeeded in facilitating policy processes in an evidence-informed way in Kenya and Bangladesh, and in supporting the development of high quality data tools to support evidence use in Pakistan and South Africa. Table 5 summarises the support provided and the headline results from each setting.

Table 5. Description of 'pilot policy' support in BCURE and the impact case study

Country	Support provided	What happened?
Bangladesh	**Support to policy processes:** BCURE provided technical support to working groups in its pilot ministries, helping them apply EIPM guidelines in a number of 'pilot policy' processes.	The working groups successfully generated policy proposals that (to varying degrees) involved robust consideration of evidence and consultation with relevant stakeholders. This helped embed knowledge and skills among EIPM trainees. The proposals were submitted to Cabinet and deemed 'noticeably different' to other policy documents by Cabinet staff. However, the first policy in the MoC led to a higher quality product and had higher levels of engagement than the pilot in the MoEF.
Kenya	**Support to policy process:** BCURE provided technical and financial support to develop a Research for Health (R4H) policy, working with a Technical Working Group made up of health research stakeholders and MoH officials.	The process produced a rigorous and comprehensive policy proposal document, achieved through a systematic process that involved sector stakeholders, and provided opportunities for three participants in BCURE training to practice and deepen their skills through conducting an evidence review. Although the R4H process was only completed to the draft proposal stage, it was felt the process would not have progressed as far without the input of BCURE, as previous attempts to establish health research priorities had stalled much earlier.
Pakistan	**Support to data visualisation tools:** including a mobile app to monitor health worker performance during polio vaccination drives; a crime map using geo-spacing mapping to identify where crime was taking place; and a tool to illustrate tax collection performance across geographical tax circles allowing managers to analyse trends.	All three pilots developed tools that could be used by front line service providers to understand what was happening on the ground and monitor performance. This contributed to decisions to adopt and in some cases scale up versions of the tools in their public sector settings.

Country	Support provided	What happened?
South Africa	**Support to data visualisation tool:** BCURE provided technical support to facilitate the production of an 'evidence map' in collaboration with the DPME: an online tool that gathered together evidence from diverse sources in relation to human settlements.	Evidence is only available from the Stage 2 evaluation, as South Africa was not included as a case study at Stage 3. The evidence map process was widely viewed as successful, helping build the capacity of DPME staff to conduct an evidence map process. Learning was documented and diffused through reports, seminars and workshops. Although this was not investigated as part of the Stage 2 evaluation, programme reports suggest that DPME successfully received funding to produce further maps in 2017.
South Africa impact case study (non-BCURE)	**Support to evaluations:** DPME provided ongoing support to government departments through specific evaluation processes, including sitting on evaluation steering committees, co-funding evaluations, advising departments on technical matters, and facilitating evaluation processes.	The case study found evidence for two broadly successful evaluation processes, which had led to finalised evaluation products and a set of recommendations that had been approved by Cabinet and had (to varying extents) been acted upon to influence the content of future policies. DPME support had also built capacity within the departments (at individual and organisational level) to conduct further evaluations.

There are a number of factors in common that explain success across the four BCURE projects and the impact case:

- **Locating a strategic entry point, where an evidence-informed approach could be showcased while also meeting priority policy objectives.** Identifying these 'win-win' situations appeared crucial to success. In Bangladesh, the more successful pilot policy benefited from a strategic choice of topic: the MoC and key external stakeholders generally agreed on the overall policy objective, and the decisions arrived at were in line with the analysis in the government's Seventh 5-Year Plan. In South Africa, the evidence map capitalised on an opportunity in the form of the Human Settlement White Paper. The policy process had been somewhat contested, with the first draft of the White Paper up for review as it was considered to be insufficiently evidence based. The need to generate a robust evidence base provided an opportunity for DPME to trial the evidence mapping methodology, which they had been interested in for some time. In Kenya, the R4H process had good ownership from the MoH stakeholders as the ministry had been pushing for a policy to better coordinate health research with national priorities for a number of years, but previous attempts had stalled. In Pakistan, the programme succeeded in identifying key problems in service delivery, in a broader context where momentum was building around data dashboards in the public sector. Finally, in the South Africa impact case, the two successful evaluations both focussed on acknowledged priority areas, with existing high-level interest and momentum. The Business Process Services evaluation had the clearest instrumental impact, largely because the Department for Trade and Industry were looking to develop the next stage of the policy being evaluated, were planning to review the success of the first phase anyway, and saw an opportunity to partner with the DPME to access resources and increase the profile of the policy.
- **Capitalising on existing work and partnerships, identifying allies, and leveraging external resources.** In Pakistan, previous and ongoing research collaborations provided a crucial entry point for the programme. For example, the tax and polio policy pilots both built on research

conducted by implementing partner Harvard – in the tax case, the research had involved digitising ten years of tax data, and the tax visualisation tool provided a way to make use of this. The polio pilot policy was successful in part because it synergised with a randomised trial conducted by the Centre for Economic Research in Pakistan (CERP) with funding from International Growth Centre (IGC), which tested the value of providing large performance-related bonuses to health workers to discourage procrastination in vaccination campaigns. This study demonstrated a significant improvement in vaccination coverage as a result of the BCURE-supported tool, which contributed to the decision to roll it out. In Kenya and South Africa, the opportunity to collaborate on the R4H policy and evidence mapping process came as a result of relationships built earlier in the BCURE programme, and the R4H process also built on a diagnostic study that identified areas where the MoH's policy processes could be improved. In South Africa, BCURE had mentored a senior stakeholder in the department, who knew the team and as a result invited BCURE to support the evidence mapping process.

- **Ensuring the right people were involved from within and outside government.** In Bangladesh, the policy process in the MoC was more successful than the process in the MoEF in large part because working group members fully bought into the process from the beginning, were present in Dhaka to attend meetings and actively participate, and had an understanding of the EIPM approach and tools (in the MoEF case, several members had not attended EIPM training beforehand). It was also important that the working groups were chaired by the 'right people' – senior enough to underscore commitment to the process, but not so senior that they were not likely to engage in any of the hands-on work. High-level support was seen as essential to the success of the process, ensuring that the activities were given priority, as they required significant time and cooperation. In South Africa and Kenya, the R4H and evidence map processes involved a mix of government staff and academic stakeholders, which helped to give the process credibility, and the R4H process was chaired by an independent and respected Kenyan academic. The South Africa impact case study suggests that an external partner can play an important coordination role between partners – DPME was asked to take the lead in coordinating the early childhood development diagnostic review given the large number of stakeholders involved, and the challenges in coordinating between them.
- **Providing flexible, responsive, tailored support through a process of 'accompaniment' and 'co-production'.** In all the cases except Pakistan, BCURE provided hands-on support to 'co-produce' policies and tools, rather than actually developing the tools themselves – although this was realised to varying degrees. This required a particular style of working widely discussed in the knowledge brokering literature, which draws on both technical and interpersonal skills. For example, in Bangladesh the national consultants were commended for their responsiveness and commitment, 'going above and beyond' to support the process, follow-up on tasks, coordinate, and 'push things along' to ensure things got done. The consultants had credibility because they possessed a practical understanding of policymaking processes and local realities, and several had previously held senior government roles. In Kenya BCURE provided evolving technical support to the unpredictable R4H process. This required developing tailored tools and inputs as the process unfolded: coordinating evidence reviews, convening stakeholder consultations and developing systematic methodologies to navigate priorities. This support was bolstered by the pre-existing trust and good working relationships established between the BCURE partners and MoH over the course of the programme, which helped overcome various obstacles. The South Africa impact case also suggested the importance of DPME's flexible, ongoing support to evaluations over an extended period of time, providing advisory support through evaluation steering committees while also ensuring departmental ownership. This role was enabled by DPME's status as a government agency with some degree of clout, given its position in the Presidency.

- **Engaging on multiple levels: building on training, and using EIPM tools to provide a structure for the process.** In Kenya, Bangladesh and South Africa, developing tools or policies through a 'co-production' approach provided opportunities for the officials involved to develop or deepen their skills in accessing, appraising and applying evidence, helping to embed the learning from EIPM training courses. In Kenya, two trainees played a key role in identifying historical gaps in national research on health systems and knowledge translation capabilities, a new insight for the MoH stakeholders. In South Africa, various participants felt they had gained understanding about the value of evidence for decision making and representation of a body of evidence, and the process helped build capacity across a core department. Finally, in both Bangladesh and Kenya, EIPM guidelines produced by the project helped to provide a stepwise, transparent structure that helped guide policy processes and ensure evidence was considered in appropriate ways at various stages. This highlights the potential of policy pilots to synergise with other programme activities – potentially a missed opportunity in Pakistan, where the policy pilots did not include individuals who had participated in the EIPM training. In the South Africa impact case, DPME also provided various support to build evaluation capacity above and beyond advisory support to evaluations – including training, developing tools and templates, and working to build capacity among evaluators on the supply side. The case study found that these activities were mutually reinforcing, helping to build capacity and buy-in for evaluation, and helping specific evaluation processes run more smoothly.

In Pakistan and South Africa, these factors helped the pilots successfully 'showcase' the value of data-driven technical tools, which led to their adoption by government agencies. Adoption was also a consequence of the fact that the tools met a recognised need and helped officials to do their jobs better. For example, the tax visualisation pilot in Pakistan allowed reports to be generated with a few clicks of a mouse that previously would have taken several days of painstaking effort, and the evidence map in South Africa was seen as a genuinely useful product that would make a practical difference to people's work.

Box 11. Insights from the literature: Showcasing and diffusion of innovations

A common assumption underpinning many BCURE interventions is that providing good examples of evidence tools or processes can 'showcase' the value of evidence, which will lead to them being adopted elsewhere. This links to 'diffusion of innovations' theory, which is about how new ideas or practices spread through imitation (Rogers, 2003). The concept has been widely applied across a range of different fields, including sociology, development studies, health, marketing and communications. Insights from development studies broadened early research on diffusion of innovations, emphasising that the meaning of an innovation may differ significantly between the agency that introduces it and the intended adopter, stressing the importance of a 'good fit' between the innovation and its potential context, and highlighting how practical demonstrations can make a potential innovation more accessible to potential users (Greenhalgh et al 2004). There are some recent examples in the EIPM literature suggesting that when external partners engage individuals in a collaborative and innovative way in the co-production of tools for EIPM, this can generate good examples that 'showcase' the value of EIPM – for example the development of an Evidence Investment Strategy in the UK that went on to inspire further strategies in other UK governmental departments, supported by interested individuals or groups (Shaxson, 2014; 2016).

In both Pakistan and South Africa, there is some evidence that ongoing government initiatives provided opportunities to scale up and roll out the EIPM tools. In South Africa, the DPME was keen to disseminate learning from the evidence map pilot through reports, seminars and workshops, and successfully sought further funding to produce evidence maps in other sectors. In Pakistan, scale-up happened in a context of a wider push within the government to modernise, including through ongoing initiatives to digitise data and use technological tools to aid service delivery. The polio pilot was rolled out in large part because it attracted the interest of the Punjab Information Technology Board, who then engaged the interest of the health board; while the tax visualisation tool inspired a much larger-scale tool with more functionality because the tax department was able to shape an ongoing digitisation process already under way within the Urban Unit. These partners (and the resources available to them) proved crucial in facilitating the adoption and scale-up of pilot initiatives. The tax visualisation case is interesting because again it highlights the value of a flexible and responsive approach, even though the pilot policies were not originally designed in this way. Senior officials were enthusiastic about the visualisation tool and wanted more from it, which was not possible to deliver through the policy pilot. However, BCURE was able to act as an interlocutor between E&T and the Urban Unit to get the functionality E&T needed into the Urban Unit's tool – something that was not part of BCURE's original Memorandum of Understanding (MoU) but which proved crucial to catalysing scale-up. Demand from senior stakeholders, generated through seeing the value of the BCURE-supported tool, was felt to be an important factor in spurring on the Urban Unit project, which had been running out of steam.

It is important to note that good quality policy products or useful data visualisation tools are only one step towards evidence actually being used to inform decision making. There are many further steps that were beyond BCURE to influence, and which are threatened by the various political and contextual barriers noted above. In Kenya, the R4H policy proposal now has to be approved by senior managers in the MoH, before going out for wider consultation, all stages that could prove highly political and will require high-level support to help the R&D unit steer the policy through to approval. There is limited evidence from Pakistan that the decision-support tools were as yet leading to more evidence-informed decision making, and there are powerful incentives for officials to make decisions based on factors other than data. Information provided by EIPM tools may challenge practices that could lead to personal gain or put an individual's role in in jeopardy by highlighting inefficiencies, which create incentives to ignore or misuse evidence.

Box 12. Summary of what worked for whom and why in 'pilot policy' activities (EQ 3)

In **Kenya**, **Bangladesh**, **Pakistan** and **South Africa**, BCURE successfully provided technical support to generate high quality policy processes or data tools (O4). These processes were most successful where policy issues were high priority or tools fit well with the priorities of and problems faced by government stakeholders (C), and where they involved a flexible and responsive approach in which the BCURE partner 'accompanied' government partners through the process (M), leading to government partners feeling a high degree of ownership and commitment to the process and its products. In **Bangladesh** and **Kenya**, projects worked through modelling a systematic approach to policy formulation (I) co-producing different inputs collaboratively (I), applying rigorous use of evidence to resolve issues within the process (I) and bringing technical and financial resources (I), leading to improvements in the quality of the resulting policy products (O6). In **Pakistan** and **South Africa**, projects developed intuitive, interactive tools (I), which genuinely facilitated officials to make decisions and do their jobs better and more easily through evidence (M). As a result, this 'showcased' the value of evidence and data for decision making (M), and enhanced government ownership of evidence tools, positioning them well for adoption by government partners (O5) who could then harness resources to roll out tools in other settings (O7).

However, uptake, use and adoption of data tools and policy proposals is challenged by political factors (C), such as the need to engage the sponsorship of senior managers to steer a policy proposal through consultation stages to final approval (C). There are also potential disincentives for officials to use tools to inform decision making where they challenge practices that could lead to personal gain, highlight inefficiencies or otherwise change the balance of power and influence in a ministry or unit (C).

Bringing the two approaches together to catalyse a step change in EIPM

While there is some evidence from Kenya, Bangladesh and Zimbabwe that BCURE contributed to improved policy processes in the short term (**O6**), there is limited evidence as yet that this has translated into support or resources to initiate further reforms to support EIPM (**O7**). Overall there is limited evidence that BCURE has promoted a step change in evidence use across target units or ministries (**impact**) – although given the short duration of the programme this is not surprising. There are some promising signs that units have been positioned to continue promoting EIPM, but whether this happens in practice depends on continued interest, opportunities and resources.

In Kenya, Bangladesh and Zimbabwe, there is some evidence that BCURE has contributed to improved policy processes in the short term. In Kenya and Bangladesh there is also some evidence to suggest the potential value of combining 'training-plus' and 'policy pilot' approaches. However, examples of policy change as a result of BCURE are mainly ad hoc, and it is too early to assess their potential impact in the long term.

- In **Kenya**, the evaluation was able to verify several cases where BCURE trainees used their skills to inform health policies, strategies and sector-wide guidelines with evidence, as well as the BCURE-supported R4H process discussed above. These included several high-profile examples that were 'given a boost' through the involvement and leadership of BCURE trainees.
- In **Zimbabwe**, there was also some evidence of trainees contributing to evidence-informed policies through the BCURE-supported research unit, although these were linked to other donor-supported programmes and it is likely that BCURE training made a relatively minor contribution.

- Finally, in **Bangladesh**, the first BCURE-supported policies are currently being reviewed by Cabinet, and the MoC has continued using the EIPM guidelines to develop further policies not supported by BCURE. However, as yet there is no evidence that BCURE has influenced policymaking in a systematic and ongoing, rather than ad hoc, way.

In all three settings, BCURE-trained officials have been able to add value to further policy products and processes in specific settings because of pre-existing incentives, policy priorities, and resources. In Kenya, the examples all occurred in units and divisions that host donor-supported health programmes, as this offered the potential for trainees to apply their skills, while trainees in other units did not have the same opportunities or lacked resources for policy formulation. Similarly in Zimbabwe, most examples of trainees applying their skills happened within donor-supported initiatives, which provided funding for research activities, and which required evidence both in support of proposals and as part of monitoring systems. These findings suggest that there would have been value in having more 'policy pilots' in Kenya and Zimbabwe, especially in high-profile policy areas supported by donor resources, both to provide more structured opportunities for trainees to hone their EIPM skills and to showcase the value of evidence.

There are some early signs that BCURE may influence future reforms and investments to embed EIPM in Kenya and Bangladesh, but limited evidence as yet. In Kenya, the evaluators noted that the R&D unit manager remains very enthusiastic about promoting evidence use in the MoH, and has integrated plans to scale up the results of the BCURE programme into the new work programme, including plans to convene a number of science cafes and develop an online evidence repository. Nevertheless, respondents felt that without a donor-funded programme, it is very difficult to implement activities at scale in the MoH and build profile and influence in the wider health sector. In the longer term, if the R4H policy is approved, this could provide an operating framework for the R4H unit to mobilise resources and catalyse investment into research for policy. But sector stakeholders suggest that this will require political support, and (again) resources, to gain traction in the MoH and wider sector. Various proposals have been put forward to donors, so far without success. There is also some evidence that BCURE played a role in catalysing further investments in Bangladesh through its work with Cabinet Division, discussed in the cross-government impact pathway.

However, overall there is limited evidence from any setting that BCURE has promoted a step change in evidence use within target units or ministries. Given the complex and politicised ministries and sectors they were working in, most BCURE projects were simply too small scale and too short to catalyse wider change over the lifetime of the programme.

In **Zimbabwe**, working with the newly established research unit in the MoYIEE was in many ways a good entry point, as the training offered an opportunity to provide technical skills to newly christened research officers in a context where other capacity building was not forthcoming. However, it is not clear that there were sufficient high-level incentives in the MoYIEE to support the embedding of EIPM after the programme left, and it is also unclear that the technical support activities were the most appropriate for catalysing organisational change. In a context where space for promoting evidence is deeply constrained by the contested and politicised nature of the youth sector, it is unreasonable to expect BCURE's relatively small investment to have contributed to a systematic shift in evidence use within the ministry in just three years.

In **Kenya**, BCURE made a little more progress in establishing capacities for evidence use in specific donor-supported programme units, and influenced some changes that could prove to be catalytic for the health sector in the longer term if funds can be mobilised to attract the necessary support. In particular, the programme made a crucial contribution to the R4H policy that would have stalled without the programme's support. However, overall there is little sign that more evidence and data is feeding through into decision making except in specific examples driven by individual practice. In Kenya, despite the support from BCURE, by 2017 the fledgling research unit was reduced to one

member of staff through retirement and promotion, so the research and evidence agenda in the MoH had stalled. In the context of the Kenyan MoH, mobilising donor funds is the means to engage senior-level and political support; without a donor-funded programme, respondents suggested that it would be difficult to attract staff to the unit as it could not offer good career incentives. These factors meant that the evidence agenda in the MoH had largely stalled once BCURE ended.

In **Pakistan** and **South Africa**, BCURE support succeeded in generating user-friendly decision-support tools, which have been adopted more widely by government agencies. It was not possible to assess the long-term influence of this in South Africa, which was not evaluated at Stage 3; and BCURE in Pakistan did not engage with the broader environment that will influence whether tools are actually used to support evidence-informed decision making. As a result, these successes represent an important but early step towards BCURE's intended impact.

Finally, in **Bangladesh**, the programme has created a solid grounding for EIPM to take root, through training large cohorts of staff and supporting ministries to develop evidence-informed policy proposals using the EIPM guidelines. Whether this catalyses into longer-term change depends to a large extent on the continued ownership and buy-in of senior staff both at ministry level, and at Cabinet Division – as discussed in the following section.

Spotlight 3. Creating spaces for conversation through networks, policy dialogues, knowledge cafes and learning events

Most BCURE programmes involved activities designed to promote dialogue and collaboration. In Kenya, Pakistan, Zimbabwe, Bangladesh and South Africa, the projects held workshops, 'knowledge cafes' or 'policy dialogues' to bring together participants from different sectors or different parts of government, such as government officials and external stakeholders (e.g. researchers, experts from industry, civil society, the media and the general public). These were generally one-off events, each involving different participants and with different topics and aims. In South Africa, BCURE housed and funded the Africa Evidence Network, a platform for professionals working in evidence production and use to engage with one another and share knowledge and resources. The ACD project also hosted international networking events through the Africa Cabinet Government Network.

There is relatively limited evidence relating to the outcomes of networking and dialogue events – this was not a core focus of the evaluation. The intended outcomes of these events were often less specific and instrumental, and more systemic and conceptual – around raising awareness and stimulating demand and momentum for EIPM across a broad network of government and non-government actors. As a result, the evaluation did not focus substantial efforts on examining these outcomes at Stage 3, given the priorities of the evaluation Steering Committee (see Annex 4).

However, the evaluation did find that where policy dialogue events were linked to specific windows of opportunity, they had the potential to shape policy processes. For example, a policy dialogue in the Ministry of Industry in Zimbabwe contributed to the revision of the Industrial Development and Trade Policies through identifying a window of opportunity (the two policies were about to expire), and convening high-profile economists and senior management in the same room to unpack the policies and provide recommendations, which were subsequently taken up in the policy review process. Similarly in Kenya, a science policy cafe on free maternal health services brought together policymakers, implementers and researchers to discuss the slow pace of the MoH's response in delivering a presidential decree on the free services policy, which generated policy actions that were subsequently taken up by the MoH through a working group.

Findings from Zimbabwe, Pakistan, Kenya and South Africa suggest that spaces for dialogue are valued in contexts where there is a recognised research-policy gap, where decision making is happening in silos, or where existing networks are dysfunctional. In Zimbabwe, knowledge café and policy dialogue events were valued because they brought together different groups to share alternative perspectives, in a highly politicised context where spaces for this type of dialogue are constrained. Such spaces can act as a rallying point around policy issues – providing small moments of discourse and engagement, creating spaces for people to discuss and understand issues, and helping to bring different actors, including researchers and policymakers, together at the same table to have debates from an a-political perspective. The project was successful in convening these events because staff created an informal environment in which people had opportunities to network and also felt able to air concerns, even on topical and heated topics; they had been strategic in using their networks to secure the attendance of high-level participants which in turn improved the credibility of events; they used knowledgeable panellists and facilitators; and over time they became increasingly more savvy in using strategic communications engage a larger audience through social media.

Box 13. Cost-effectiveness of policy dialogues and knowledge cafes

The evaluation found that in the countries where these activities were used and where we had data (Kenya and Zimbabwe) the costs of running knowledge cafes and learning events was relatively low: between £8,000–10,000 per event and the cost per participant of around £220. This suggests that if these events are used strategically and opportunistically, they can prove highly cost-effective.

6. Impact pathway 2: Cross-government approach

Three BCURE projects worked to promote EIPM across governments, rather than (or alongside) work within specific ministries. In Bangladesh, BCURE worked with Cabinet Division to develop and roll out government-wide EIPM guidelines, and pilot them in three line ministries. In Pakistan, an EIPM training course was rolled out across the whole civil service. Both of these cases were examined in depth in the final stage of the evaluation. Further insights are drawn from the Stage 2 evaluation in Sierra Leone, and the non-BCURE impact case study of the South Africa National Evaluation System.

Box 14. Summary of BCURE cross-government support

Three approaches: 'top down', 'bottom up' and 'institutionalising training'

In **Bangladesh**, BCURE promoted EIPM simultaneously from the 'top down' and the 'bottom up'. From the 'top down', the project worked with Cabinet Division to develop and roll out EIPM guidelines, which sought to establish common cross-government standards for EIPM. The guidelines present a five-step approach to policy formulation: from problem definition, through objective setting, identification of policy options and impact assessment, to comparison and selection of the 'best' policy option. BCURE also facilitated various learning events, exchanges and training with Cabinet officials to build high-level buy-in for EIPM, and worked with (mainly senior) officials in line ministries to facilitate them to develop pilot policies using the guidelines. From the 'bottom up', the programme developed an EIPM training course and delivered it to around 400 civil servants responsible for policy formulation in three line ministries. BCURE also worked to institutionalise EIPM training, embedding EIPM curricula within two national public sector training institutes.

In **Pakistan**, BCURE primarily worked from the 'bottom up', through delivering EIPM training on a large scale, consisting of online modules and in-class lectures. This course was rolled out to 1,780 civil servants through the National School of Public Policy (NSPP), who has institutionalised the training into the national curriculum for mid-career officials, senior management and national management.

In **Sierra Leone**, BCURE primarily promoted EIPM from the 'top down' through introducing a comprehensive manual of Cabinet procedures. This included templates for ministries to make policy submissions to Cabinet which required consideration of evidence. The use of evidence is just one part of these new procedures, which also focus on better planning, coordination with other ministries, efficiency in Cabinet through more structured discussion, and accountability. The programme provided policy analysis training to Cabinet Secretariat staff, established and revised structures in the Secretariat including a Cabinet Policy Review Unit and a Cabinet Implementation Monitoring and Support Unit, and hosted international roundtable events for Cabinet Secretaries from participating countries to network and interact. It also created and trained a Cabinet focal person (CFP) network in line ministries to liaise between the ministry and the Secretariat. In its final year, the project expanded to encompass a 'bottom up' approach through training 150 policy staff from line ministries; but this was not investigated in the Stage 3 evaluation.

Non-BCURE case study of the South Africa National Evaluation System

In **South Africa**, the DPME provides 'top down' support to the National Evaluation System (NES), which was established in 2012 to promote and support evaluations across government. DPME produces an annual National Evaluation Plan, works to stimulate demand for evaluations by senior government managers, supports departments to commission and conduct evaluations (discussed under the single ministry impact pathway), and promotes high quality evaluation practice through producing guidelines, templates and standards. As of 2017, 69 evaluations were completed or under way under the NES.

6.1 What were the drivers, opportunities and risks for EIPM in the cross-government settings?

Bangladesh, Pakistan and Sierra Leone present challenging political environments for promoting EIPM. However, in all three countries there were opportunities to implement cross-government reform through high-level institutional champions.

Bangladesh and Pakistan both score poorly on indicators of governance and have records of high corruption. Both countries face similar challenges in their civil services, with limited resources and overstretched systems, periodic mismanagement of finances and corruption, technical capacity constraints around policy formulation and implementation at national and provincial levels, and a trend for ongoing politicisation of the civil service.

However, in Bangladesh, governance reform has been on the agenda for decades, supported by donor partnerships with the World Bank, UNDP and DFID among others. Various reforms have enhanced emphasis on evidence, including the introduction of a new Budget Framework and increasing reliance in government on standard operating procedures. One of the most powerful regulatory institutions, Cabinet Division, has gradually taken on the role of coordinating and championing reform, and plays a central role in policy coordination among the ministries and divisions. Cabinet Division provided a key entry point for BCURE, given the alignment between these reform priorities and BCURE's objectives, with the initial relationship facilitated by the support and enthusiasm of the then-Cabinet Secretary. Programme activities were housed in the Coordination and Reforms Wing, a new entity at the time of project start up with a mandate to advance various government reforms. As a senior Cabinet stakeholder explained: "*We already wanted this sort of project, but we didn't find any donors [before BCURE]*."

In Pakistan, BCURE had an opportunity to promote large-scale reform through pre-existing relationships between implementing partner Harvard and the National School of Public Policy (NSPP), which has a mandate to improve the quality and effectiveness of the public service in Pakistan, and trains every federal, provincial and district level civil servant multiple times throughout their careers. Partnering with the NSPP presented a strategic opportunity to embed EIPM training across the whole civil service, in the context of the civil service reform framework. Harvard formed an early link with the Rector who heads the NSPP, a key champion for the programme and for the later expansion of the training to different cadres of civil servants.

Sierra Leone is a fragile state, facing the challenge of weak government capacity and lack of trust in government institutions. The Ebola crisis from 2014–16 led to a state of emergency being declared, causing major disruption to government development plans. However, the country has a functioning Cabinet system, with most major executive decisions taken by Cabinet. In 2013 a new Cabinet Secretary was appointed, a position that the Sierra Leone Constitution combines with the head of the civil service into one role. This individual proved to be an important champion for EIPM, providing an entry point for the programme to institute broad-based reforms.

The governance challenges in the BCURE contexts raise questions around how far EIPM tools will be meaningfully institutionalised, and how far trainees will have opportunities to apply new skills. In Pakistan, the evaluation found that there are few existing processes to support evidence access and use at an organisational level in the civil service, and general inexperience in drawing on evidence in decision making including among trainees' managers. This creates a challenging context for trainees to apply new skills while b6ack in the workplace. In both Bangladesh and Sierra Leone, systemic weaknesses in government institutions and bureaucratic dependence on donors to support policy processes raise the potential risk of isomorphic mimicry, where new EIPM procedures may be adopted on the surface without leading to real change in day-to-day practice at line ministry level (see Box 3 above).

6.2 What happened and why?

Figure 5 depicts how BCURE worked towards a step change in evidence use through providing government-wide support. It does not represent what any individual project did, but rather synthesises evidence from across BCURE and the broader literature on how and why capacity support can lead to change. The diagram summarises what the evaluation has learned about how capacity support can contribute to EIPM in a cross-government setting, but it is not a fully tested theory, as projects made different degrees of progress towards the intended impact. It is therefore intended as a broad 'road map' for future programmes working to promote EIPM across governments, rather than a definitive prescription.

Below the diagram Table 6 presents an overview of how far each of the outcomes were achieved by the BCURE projects (EQ 1) and BCURE's contribution (EQ 2). These findings are unpacked throughout the rest of this section, which also explores how and why BCURE did or did not make a difference (EQ 3).

Figure 5. Cross-government impact pathway

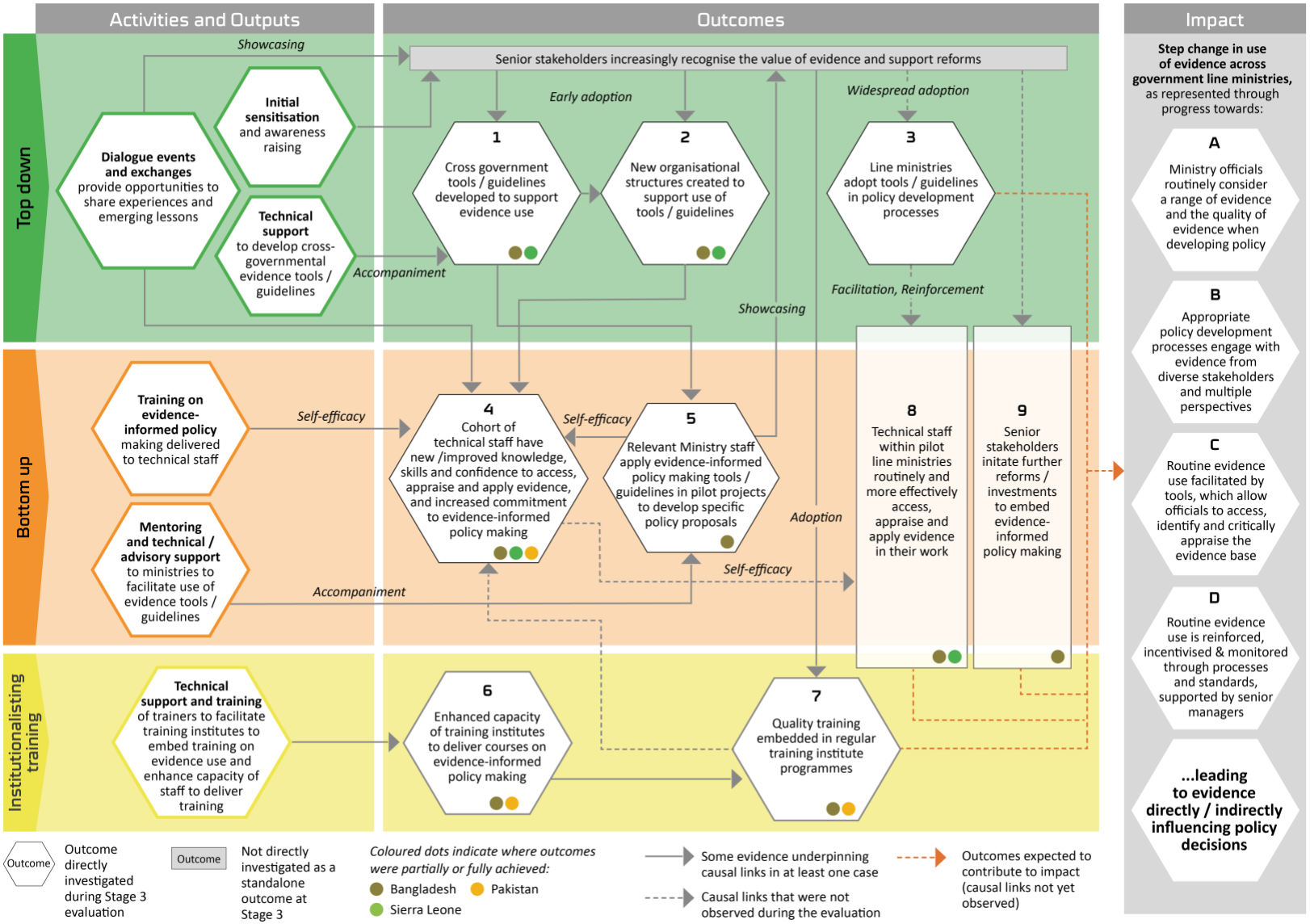


Table 6. Summary of evidence for cross-government impact pathway

Outcome	Summary of evidence for outcome (EQ 1) and BCURE contribution (EQ 2)
'Top down' approach: Bangladesh and Sierra Leone	
1. Cross-government tools or guidelines developed to support evidence use	Strong evidence in Bangladesh and the Stage 2 evaluation of Sierra Leone that BCURE directly facilitated the creation of new tools in the form of guidelines (Bangladesh) and Cabinet procedures and templates (Sierra Leone) to support EIPM, in close collaboration with Cabinet Division/the Cabinet Secretariat.
2. New organisational structures created to support use of tools or guidelines	Strong evidence from the Stage 2 evaluation in Sierra Leone that BCURE helped establish permanent new organisational structures in cabinet and line ministries to support the use of new cabinet procedures, including a cabinet focal person (CFP) network in line ministries to liaise between ministries and the Secretariat. However, it was not possible to assess the status of these structures at the end of the programme as this country was not included in the Stage 3 evaluation. In Bangladesh, BCURE facilitated the establishment of working groups in line ministries, to support the use of EIPM guidelines in pilot policies, but these were not intended to be permanent structures.
3. Line ministries adopt tools or guidelines in policy development processes	Some evidence from the Stage 2 evaluation in Sierra Leone that new cabinet procedures were adopted at line ministry level, in that CFPs had been established in line ministries and ministers were aware of the procedure and the need to use it. However, the extent and depth of this adoption was unclear and it was not possible to assess the status of the procedures in line ministries at the end of the programme. The guidelines had not yet been formally circulated to line ministries by Cabinet Division in Bangladesh at the time of the final evaluation. Based on the views of line ministry staff, it seemed likely that some level of adoption would occur, but it was too early to assess how far this would be meaningful and sustained.
'Bottom up' approach: Bangladesh and Pakistan	
4. Cohort of technical staff have new/ improved knowledge, skills and confidence to access, appraise and apply evidence, and increased commitment to EIPM	Strong evidence from Bangladesh and Pakistan that the EIPM training courses made a crucial contribution to increasing the knowledge, skills and attitudes of trainees in relation to understanding how to access, use and appraise evidence to inform policy decisions. Some evidence from Bangladesh that the BCURE-supported pilot policies also provided opportunities for technical staff to apply and deepen skills developed through the training – this was not the case in Pakistan, as trainees were not involved in the pilot policies discussed under the single ministry pathway (this was not part of the project design).
5. Relevant ministry staff apply EIPM tools or guidelines in pilot projects to develop specific policy proposals	Strong evidence from Bangladesh that BCURE had directly facilitated relevant ministry staff to apply EIPM guidelines in the development of pilot policies. As above, this was not part of the cross-government approach in Pakistan.

Institutionalising training: Bangladesh and Pakistan	
6. Enhanced capacity of training institute trainers to deliver EIPM courses	Strong evidence from **Bangladesh** and **Pakistan** that BCURE made an important contribution to enhancing the capacity of local training institute trainers to deliver EIPM courses, although in Pakistan some trainers had already moved to different roles by the end of the project, and there were no concrete plans in place to continue replenishing the training pool.
7. Quality EIPM training embedded in regular training institute programmes	Strong evidence from **Bangladesh** and **Pakistan** that BCURE made an important contribution to embedding EIPM training into the ongoing programmes of national training institutes. However, there are concerns in Pakistan that the model is not sustainable because Harvard has ongoing hosting costs associated with the online training platform and there is not yet funding in place to pay for continued access.
Longer-term change across all three approaches	
8. Technical staff within pilot line ministries routinely access, appraise and apply evidence in their work	Limited evidence from **Bangladesh**, **Pakistan** or **Sierra Leone** (at Stage 2) that BCURE had led to widespread shifts in evidence use among technical staff in line ministries. In **Bangladesh**, there is strong evidence that MoC staff have continued to use the BCURE-supported EIPM guidelines in a small number of policy processes beyond the initial policy pilot identified by the ministry working group, but there was limited evidence that this had happened in the MoEF, and also limited evidence that trainees not involved in policy pilots had used the training in their work (although some had only recently finished the training). In **Sierra Leone**, there was some evidence at Stage 2 that new cabinet procedures were being used by line ministries, but it was not possible to assess how sustainable this was as this country was not included in the Stage 3 evaluation. In Pakistan, examples of evidence use by trainees generally represented isolated pockets of evidence use, driven by individuals, rather than more routine use of evidence within teams or divisions – although the evaluation was only able to interview a small proportion of trainees.
9. Senior stakeholders initiate further reforms and investments to embed EIPM	Some evidence from Bangladesh that Cabinet Division was spearheading further EIPM reforms and investments, including through directing the ten largest spending ministries to establish Policy Research Units, and that BCURE had made an important contribution to this decision. However, it is too early to say what might come of this initiative, and how it will ultimately affect evidence use in policymaking. There is no evidence for this outcome in Pakistan (in relation to its cross-government work as opposed to its support to policy pilots) or Sierra Leone.
Impact: Step change in use of evidence across government	In Bangladesh, BCURE has made good progress towards the longer-term impact through combining 'bottom up' support to individual skills with 'top down' strengthening of the government's institutional framework for EIPM – although continued progress depends on the ongoing engagement of Cabinet Division and there are a number of risks to this. In Sierra Leone, it is not possible to make a final assessment of progress towards the impact as the latest evaluation data was collected in 2016, several months prior to programme completion. However, there was some evidence that important progress had been made at Stage 2, through establishing new systems, structures and procedures that provided a strong foundation for EIPM by demanding evidence in proposals to Cabinet. There is no evidence that BCURE in Pakistan has made much progress towards a step change in the use of evidence across government. A number of ad hoc examples of policy-level change were observed where trained individuals were able to apply their skills to improve evidence use within specific policy processes as part of their day-to-day work. However, there was no evidence that these ad hoc examples were catalysing more routine evidence use across government.

'Top down' approach

In Bangladesh and Sierra Leone, BCURE facilitated the creation of new templates and guidelines to support EIPM (**O1**). In Sierra Leone, BCURE also helped establish permanent new organisational structures in cabinet and line ministries to support their use (**O2**).

In Bangladesh and Sierra Leone, 'top down' EIPM tools and guidelines were successfully developed and finalised as a result of clear government ownership, and a sensitive and flexible implementation approach. In Bangladesh, BCURE facilitated the development of EIPM guidelines in close collaboration with Cabinet Division. These were finalised following piloting at line ministry level, and at the time of the final evaluation visit were close to being formally endorsed by the government. In Sierra Leone, the programme succeeded in establishing new cabinet procedures and policy templates that required consideration of evidence to support proposals to Cabinet. This was described as "*the greatest change in Cabinet decision-making procedures since our independence in 1961."* In both countries, the new guidelines or procedures were adopted at cabinet level because:

- **Cabinet Division in Bangladesh and the Cabinet Secretariat in Sierra Leone had strong ownership over the process, deriving from a mandate for reform and improvement that clearly aligned with BCURE's objectives, and leadership by a high-level champion.** In Bangladesh, there was a strong alignment between BCURE's aims and the government's administrative and governance reform agenda. Importantly, there was a high-level champion in the form of the Secretary of Cabinet Division at the time of the project's inception, who was personally very interested in the programme and clear what he wanted from it, and helped bring other high-level stakeholders on board. In Sierra Leone, the Cabinet Secretary also played a key role in championing the development and roll-out of the new procedures, and was described as highly committed to bringing change and improving transparency in decision making. These experiences highlight the importance of committed individuals who can act as gatekeepers and cheerleaders for a project. However, in Bangladesh the Secretary changed roles during BCURE, which created problems in terms of maintaining momentum, and highlights the risks of over-dependence on charismatic individuals.

  In the South Africa impact case, respondents emphasised the importance of the DPME in driving the NES, providing it with an institutional home. The NES had emerged as a consequence of high-level demand for improved government performance. As in Sierra Leone and Bangladesh, this suggests that pressure for performance improvement spurred demand for evidence and created the momentum to invest in new systems.

- **In both Bangladesh and Sierra Leone, the implementation approach was sensitive, flexible, tailored to the local context and encouraged ownership.** This proved crucial in gaining and maintaining senior-level buy-in. In Bangladesh, the programme benefited from a local partner (PRI) who had access to senior stakeholders and a deep understanding of the context, and an implementing team who had the interpersonal skills to engage successfully with senior government stakeholders, and the willingness (and savviness) to 'walk at the pace of the government'. PRI advice in the early stages of the programme resulted in a number of changes that promoted greater Cabinet Division ownership, which appeared crucial to the programme's success. This also appeared important in Sierra Leone, where the revised manual was developed through a highly consultative, participatory and iterative process involving inputs from all Cabinet members and ministries. The Cabinet Secretary played a key role in facilitating discussion among government stakeholders during the revision of procedures, allowing sufficient time for stakeholders to come on board but also using political capital to nudge and drive forward the process while managing resistance. In South Africa, the impact case study found that ministry engagement in the NES was enabled by the technical competence of the DPME and the team's

ability to convene stakeholders, promote trust and work collaboratively with line ministries through evaluation steering committees.

- **In Bangladesh, framing EIPM as a technical approach to improve the quality of policy may have also attracted government stakeholders.** This view was echoed by many officials across Bangladesh's Cabinet Division and line ministries, who perceived the EIPM guidelines as providing a structured approach to policy formulation that could strengthen the quality of policy products – essentially helping to facilitate staff efforts to formulate policies more effectively and subsequently get them approved. In Sierra Leone too, the absence of existing systematic policy development processes presented an entry point for BCURE through improving policymaking procedures more broadly, in a way that promoted evidence access, appraisal and use.

After being approved by Cabinet, new EIPM tools were adopted in Sierra Leone at line ministry level (**O3**), in part due to the various structures in place to support implementation (**O2**). The EIPM guidelines seem likely to be adopted in Bangladesh following an instruction from Cabinet Division, helped by the generally positive experience of policy pilots which 'showcased' the value of the approach to senior ministry staff (**O5**) – but there is limited evidence from either Bangladesh or Sierra Leone on how meaningful adoption is or will be.

The EIPM guidelines had not yet been formally circulated by Cabinet Division in Bangladesh at the time of the final evaluation, so it was too early to assess line ministries' reaction. However, based on the views of line ministry staff, it seemed likely that some level of adoption would occur. Many junior and senior staff from both pilot ministries indicated that once the guidelines were formally approved by Cabinet Division they would be viewed as mandatory at a ministry level. However, there are many rules and guidelines in the Government of Bangladesh already, and comments by some senior staff suggested a degree of uncertainty around how far new rules will be taken on board despite the directive from Cabinet Division – which in turn will affect how technical staff use or do not use the guidelines (and their training) in their work. What seems key is maintaining the buy-in and commitment of high-level bureaucrats for the guidelines to be meaningfully adopted and sustained (discussed further below). There are serious risks to this in a context of regular senior civil servant turnover in Bangladesh – the programme has been affected by multiple senior champions changing jobs and being replaced by new individuals without any knowledge of the programme or the EIPM agenda.

There is some suggestion – although the evidence is fairly limited – that the BCURE-supported policy pilots in Bangladesh helped to 'showcase' the value that evidence can bring to policy formulation processes, building senior-level buy-in within line ministries that may help support future adoption. BCURE ensured the pilot policy working groups consulted ministry secretaries, who chaired the process and were consulted at each stage – although in practice the level of involvement depended on the relationship between the head of the working group and the secretary. Some staff suggested that the fact that the first three policy pilots were more or less 'successful' was important to demonstrating the value of EIPM, including the guidelines, to both Cabinet Division and senior ministry staff, giving them greater confidence in the approach. The Ministry of Commerce had continued applying the approach informally in four further policy processes without hands-on support by BCURE, suggesting that the initial pilot process did have some positive demonstration effect. However, some stakeholders were sceptical that the relatively small-scale pilot policies would be sufficient to 'showcase' the value of the EIPM process across a whole ministry or the Government of Bangladesh.

In the South Africa impact case, the DPME took a similar approach of involving high-level ministry stakeholders closely in their evaluations, from commissioning through to taking decisions in response to the findings, which provided senior stakeholders with a practical demonstration of the value of using evidence and helped build commitment to the evaluation agenda.

In Sierra Leone, there was some evidence from the Stage 2 evaluation (conducted in 2016, a few months prior to the end of the programme) that the revised Cabinet memo had been adopted by line ministries to some extent, but there was still a 'long way to go' in terms of implementation. BCURE had established and strengthened various organisational structures both at Cabinet and line ministry level (described in Box 14 above) – this programme feature was not present in either Bangladesh or Pakistan. There was some evidence that this had helped support the implementation of the new procedures, through both 'carrots' (the newly established CFPs in line ministries providing practical support in implementing the new procedures, who were in turn supported by the new Policy Review Unit in Cabinet) and 'sticks' (the Cabinet Implementation Monitoring and Support Unit closely following the implementation of proposals). Officials reported two or three cases where ministers had insisted on submitting proposals using the new procedures without taking the advice of the Cabinet Policy Review Unit, and the proposal had been thrown out. Establishing the Cabinet Implementation Monitoring and Support Unit also increased pressure on ministries to make proposals realistic, given that implementation will now be followed up closely. These top-down mechanisms are important because the CFPs have limited power to push back if a minister insists on a policy that is insufficiently evidence-informed going ahead.

The South Africa impact case found a similar reliance on both carrots and sticks: on the one hand funding and hands-on support from the DPME to commission and oversee high quality evaluations, and on the other improvement plans that must be submitted to Cabinet alongside the evaluations and regularly reported on, which helps ensure ministries act on evaluation recommendations.

In Bangladesh it is unclear as yet if and how Cabinet Division will continue supporting and enforcing use of the guidelines now that the programme has ended, and programme staff felt there is an issue of 'missing incentives' at ministry level to make evidence-informed decisions which BCURE was unable to address. The short duration of the project also poses a risk to sustainability, as it means an end to the ministry-level support the programme offered, which has been crucial to the use of guidelines in pilot policy processes.

Compliance with procedures and templates will also not necessarily lead to improvements in evidence use. In Sierra Leone, staff who reviewed submissions reported improvement in the presentation of existing evidence, but this did not necessarily signify a change in how evidence was actually used to guide decision making, and as yet there were limited observed changes in the quality of evidence considered in proposals. While line ministries emphasised a number of changes they had made so far to their policymaking practice as a result of the revised procedures, aspects other than evidence use seemed to be more prominent. There were many factors that might inhibit ministries from actively engaging with evidence despite the adoption of the new procedures, including the low quality of data available to develop proposals, lack of funding and resources for basic facilitates in ministries (including internet and transport to support data collection), and an organisational culture slow to change, characterised by some participants as a 'rush to wait.'

The South Africa impact case study provides some clear lessons about 'what happens next' following the formal adoption of an EIPM tool or process, given that the National Evaluation System was set up in 2012. As in Bangladesh and Sierra Leone, the case study emphasised the importance of the DPME having a clear mandate to engage with ministries to promote EIPM. However, the successes observed through this case study did not derive from the DPME simply directing line ministries to conduct evaluations, but through their ongoing work in driving the evaluation agenda in South Africa from multiple angles – including through maintaining momentum within evaluation steering committees, engaging with evaluators to build supply-side capacity to meet demand, and demonstrating the value of evaluation and evidence for improving performance. An ongoing evaluation of the NES has found that ministry engagement with evaluation has been uneven – as despite DPME's mandate and position of influence within the Presidency, and the co-funding model that incentivises departments to conduct evaluations, decisions to engage with the

NES are voluntary. There are now plans to develop legislation for evaluation in order to further strengthen DPME's mandate and emphasise that evaluations must be undertaken and recommendations followed up.

Box 15. Summary of what worked for who and why in establishing 'top down' cross-government tools and structures (EQ 3)

In **Bangladesh** and **Sierra Leone**, Cabinet Division and the Cabinet Secretariat decided to adopt and endorse (M) new EIPM tools and systems (O1) because these institutions had clear ownership over and buy-in to the process (C), in part a consequence of the support of high-level champions (C), and in part because they had a mandate for reform aligned with BCURE's objectives (C). In Bangladesh, the framing of EIPM as a technical approach to improve policy formulation was a key selling point (I). Ownership was also promoted through BCURE's implementation approach, which was sensitive, flexible, and tailored to the local context (I): an approach that can be characterised as 'accompaniment' (M).

Following a high-level directive from Cabinet Division / Secretariat (C), EIPM tools and guidelines were adopted by line ministries in **Sierra Leone**, and seem likely to be adopted in **Bangladesh** (O3), where the policy pilots have to some extent successfully showcased (M) their value. However, it is too early to determine whether this will lead to routine use of evidence by line ministry staff (O8). Insights from **Sierra Leone** and the **South Africa** impact case suggest that a one-off directive is not enough: ongoing engagement through both 'carrots and sticks' (C), enforced by a senior institutional champion (C) is necessary to ensure new tools and processes actually change behaviour (O4). In **Bangladesh**, there is a risk that there are insufficient incentives at ministry level to catalyse senior demand (C), that senior staff who have shown interest in the approach will move on (C), and that ministries will struggle to implement the guidelines without the ongoing support of the programme, in part a consequence of its very short implementation period (I).

'Bottom up' approach

In both Bangladesh and Pakistan, BCURE has delivered EIPM training to a large cohort of civil servants. This has increased knowledge, skills and attitudes around evidence access, appraisal and use (**O4**) but as yet there is limited evidence of widespread or routine behaviour change (**O8**).

In both Bangladesh and Pakistan, there is strong evidence that the EIPM training courses increased the knowledge, skills and attitudes of trainees in relation to evidence use, but limited evidence of widespread or routine behaviour change among trainees. Programme M&E systems were not set up to monitor behaviour change across whole cohorts of trainees, and it was only possible for the evaluation to interview a small proportion of the large numbers trained – therefore the evidence on these outcomes is relatively limited. However, in both Bangladesh and Pakistan, interviews suggested that many trainees had been unable to apply their new knowledge and skills in their work as yet.

In Bangladesh, a small number of trainees and other ministry staff were supported to apply their learning and deepen their skills through 'pilot policies' using the EIPM guidelines, which helped build individual-level skills (discussed in the single ministry impact pathway) as well as showcasing the value of an evidence-informed approach (discussed above). However, most trainees had not yet had an opportunity to apply the training, and while some had only recently attended the course, others

felt there were limited opportunities in their roles to apply what they had learned. Several stakeholders suggested that without follow-up (for example refresher training, or permanent EIPM focal points within ministries) there is a risk that trainees will forget what they learned or will lack the confidence to apply their learning, which resonates with findings from across BCURE as well as the wider literature on adult learning discussed in Spotlight 2. The hope is that the programme has managed to catalyse sufficient top down buy-in to EIPM within Cabinet Division and line ministries, which will create incentives for trainees in relevant positions to apply the concepts learned through the training – but it is too early to test this theory.

In Pakistan, where interviewed trainees had been able to apply their learning, this generally represented isolated pockets of evidence use in specific tasks, where training had helped officials think about how to use data differently. There were a number of common contextual factors at play where trainees had been able to apply their skills. On an individual level, trainees often had high levels of prior education, good soft skills enabling them to influence senior managers to shape a policy or process, and high levels of self-initiative and drive. On an organisational level, trainees had been presented with an opportunity to use evidence in relation to a specific task, had supportive senior management, and were allocated or able to draw on research to access and analyse data. However, many trainees interviewed for the evaluation felt the training was not directly relevant to their professional roles and so there would be limited opportunities to apply their learning. This was recognised by programme staff as a necessary consequence of a training model that targets all civil servants at particular levels as part of mandatory procedures, rather than targeting specific staff based on the relevance of training content to their roles. Trainees also indicated that there were missing incentives in their workplaces to change practices towards more evidence-informed policymaking – linked to uninterested senior managers themselves lacking an incentive to consider evidence, and corruption providing a motive to ignore or suppress evidence.

The training approach in Pakistan was premised on a theory that training a large cohort of officials would help lead to institutionalisation of EIPM through a 'critical mass' effect; but evidence from across BCURE suggests this is unlikely to work without top down reform. Underpinning the mass training approach in Pakistan was the assumption that training large numbers of civil servants across government would lead to improved knowledge, skills and attitudes about the importance of EIPM among a broad cohort of officials, which would eventually catalyse a shift in practice. The BCURE programme in Pakistan identified that attitude change was a key first step towards promoting evidence use. Programme staff suggested that the critical mass effect involved embedding new ideas and attitutdes in the minds of large numbers of civil servants, which, when undertaken over a long-term period of five to ten years, should lay the groundwork for a 'culture of evidence use' to emerge:

> *We recognise that one-off training is not going to transform the way things are done but it is an important place to at least introduce ideas and concepts they have not encountered before through the approach that is targeting civil servants all levels… It will take some time to achieve systematic change.*
> *(Programme staff member, Pakistan)*

At Stage 2, the evaluation attempted to interrogate how this 'critical mass' effect might work, and unpacked it into different theories (see Box 9 above). Given the need for large numbers of officials to be trained, over a long-time frame, it may be too early in BCURE to fully test the 'critical mass' theory. However, the evidence from Pakistan suggests that there are risks to relying solely on this strategy. Although there was evidence that the training had influenced more positive attitudes towards the usefulness of data and evidence, there was no evidence that this had gone beyond raising individuals' awareness and, for some, contributing to ad hoc instrumental change in individuals' practice. Evidence from across the BCURE programme as well as the wider literature on

adult learning (discussed in Spotlight 2) suggests that, while attitudinal change is important, systemic change is unlikely to come about through mass training alone without addressing some of the incentives and organisational structures that inhibit evidence access, appraisal and use in policymaking. Trainees are likely to forget their learning if they do not have opportunities to apply their skills straight away, are not supported through follow-up activities, or required to apply their skills through top down demands. Insights from Kenya and Zimbabwe, discussed in the single ministry impact pathway, also suggest that a critical mass is less likely to be reached without an effort to cluster trainees in order to catalyse pockets of good practice, or engage trainees' managers in order to stimulate support and demand to help them apply their learning.

Box 16. Summary of what worked for who and why in promoting cross-government reform through large-scale training (EQ 3)

The implicit theory in **Pakistan** was that training many civil servants across government (I) would lead to a broad cohort of trained officials with awareness of EIPM ideas and more positive attitudes towards the role of evidence in policy making. Over time, this would contribute to a 'critical mass' of people working differently (M), which would catalyse a broader shift towards a culture of evidence use in the civil service (O8). It is too early to decisively reject this theory, but as yet the evidence suggests that, while they may have a more favourable attitude towards evidence use and improved skills, many trainees have been unable to apply their learning or influence the practice of colleagues or managers. Evidence from across the BCURE portfolio and the wider literature also suggests that achieving a critical mass is unlikely without addressing the incentives and organisational structures that inhibit evidence access, appraisal and use (I), considering how to cluster trainees to generate pockets of changed practice (I), or engaging senior managers to stimulate support for trainees to apply their learning (I).

In **Bangladesh** the hope is that Cabinet Division will continue to use its clout to promote the EIPM agenda (C), leading to senior staff in line ministries reinforcing better use of evidence within policy development processes through making increased demands of staff (M), which staff will be able to deliver as a result of increased self-efficacy through training (M) and EIPM guidelines acting as a facilitator (M). However, it is too early to determine whether this theory will be borne out in practice, and there is limited evidence as yet of routine change in practice (O8).

'Institutionalising training' approach

In both Bangladesh and Pakistan, BCURE has succeeded in building the capacity of local training institutes to continue delivering EIPM training (**O6**), and institutionalising EIPM courses in national civil service training centres (**O7**). However, this alone is unlikely to lead to widespread shifts in evidence use across a civil service.

There is strong evidence in Bangladesh that BCURE has strengthened the capacity of two key national training institutes to deliver EIPM training, and helped embed training into their ongoing programmes. In one training institute alone, upwards of 1,600 civil servants will be exposed to EIPM annually through its programmes. This institutionalisation was possible given Bangladesh's strong culture of civil service training, and the presence of existing courses into which EIPM content could be embedded. Importantly, EIPM was adopted into training institute curricula following Cabinet Division instructions to consider the EIPM training manual in their policy-related training courses. Programme staff also indicated that some of the training institute stakeholders have the potential to

be high-level champions for EIPM, who can potentially 'keep momentum going' after the end of the programme.

In Pakistan, the institutionalisation of training was a core part of the programme model from the outset, and there is strong evidence that BCURE succeeded in building local capacity to deliver EIPM courses. The adoption of EIPM training into national training courses for civil servants was possible because there was a clear 'home' in the form of the National School of Public Policy (NSPP), which is the single entity that delivers training to all civil servants nationally. A crucial factor behind this adoption was the fact that the NSPP Rector at the time of the programme inception fully bought in to BCURE's suggested approach, and became a vocal supporter of EIPM. This high-level support for the training percolated to lower levels of senior management, who pushed for the BCURE training to be brought to different cohorts of trainees and applied more widely. Harvard actively engaged with the NSPP throughout the programme to build and maintain this buy-in – and importantly Harvard had a relationship with NSPP pre-dating BCURE, which provided an entry point for the institutionalisation of the training course. However, there are some concerns in Pakistan about the sustainability of the training model and the ongoing use of the online training platform as there is not yet funding in place to pay for continued access (at a relatively modest cost of US$200 per person to cover hosting and maintenance costs). Nevertheless, there are indications that donor funds are only required for a further two years before the NSPP is able to fully support the training with its own resources, and in the meantime, NSPP and the Harvard team are working closely together to ensure the sustainability of the programme.

While institutionalising training is an important step towards ensuring the sustainability of BCURE projects, on its own it is unlikely to catalyse a step change in evidence use. As summarised in Box 16 above, training alone is not enough to promote widespread behaviour change. There is also a risk that courses embedded in national training institutes will be less intensive than BCURE-supported training – for example in Bangladesh the 5-day BCURE course has been considerably shortened to incorporate it into existing courses. Where training is incorporated into existing curricula this may also lead to it becoming much more broad-brush, rather than targeting individuals who are likely to have specific opportunities to apply their skills, as observed in Pakistan where training is mandatory for all staff at a particular level.

Box 17. Summary of what worked for who and why in institutionalising training (EQ 3)

EIPM training was successfully adopted (M) into national training curricula in both **Bangladesh** and **Pakistan**, and local trainers capacitated to deliver the courses (O). This was possible in a context where there was an established culture of civil service training, strong training colleges with existing courses that could be modified to include EIPM, and a national training pool who already had a good level of training capacity (C). In **Pakistan**, pre-existing relationships between the programme and the national training provider, plus the enthusiasm of a high-level champion, helped create an entry point for institutionalisation (C) and expanding the course to new cohorts (O). In **Bangladesh**, the course was institutionalised due to the support of Cabinet Division built up over the course of the project, who had the power to direct the training institutes to embed the EIPM course (C).

Bringing the three approaches together to catalyse a step change in EIPM

BCURE made most progress towards a step change in evidence use across governments (**impact**) in Bangladesh and Sierra Leone, where the programme design involved establishing 'top down' structures for EIPM supported by strong Cabinet ownership. More limited progress was noted in Pakistan where the programme did not take this approach.

In Bangladesh, BCURE made substantial progress towards the longer-term impact, and contributed towards senior stakeholders initiating further EIPM reforms as a result of its systemic approach combining 'bottom up' support to individual skills with 'top down' strengthening of the institutional framework for EIPM. Further progress towards a step change in evidence use will depend on the extent to which EIPM is meaningfully adopted by line ministries following its endorsement by Cabinet Division, and promoted by senior line ministry staff. The ongoing ownership of Cabinet Division is also important – including how far it is prepared to take further steps to invest in supporting evidence use. Insights from Sierra Leone and the South Africa impact case suggest that adoption will require more than a one-off instruction to use the new approach. It will require Cabinet Division to continue to employ both 'carrots' and 'sticks' to encourage ongoing engagement with EIPM by line ministries.

There are some promising early signs in this regard. The final evaluation found that Cabinet Division in Bangladesh had directed the ten largest spending ministries to establish Policy Research Units, and that BCURE had made an important contribution to this decision. The idea was that these units would support research work, including collection and analysis of evidence. High-level government champions in the BCURE project Steering Committee had played an important role in promoting the Policy Research Unit idea, in pushing for institutionalisation of BCURE-associated reforms. This unintended outcome is a positive sign that the programme has succeeded in getting Cabinet Division and other high-level stakeholder to genuinely buy into EIPM, although it is too early to say what might come of this initiative.

In Sierra Leone, it is not possible to make a final assessment of progress towards the impact as the latest evaluation data was collected in 2016, several months prior to programme completion. However, important progress was noted at Stage 2. Evidence from Stage 2 suggested that the programme had made significant progress towards the impact through establishing new systems, structures and procedures that provided a strong foundation for EIPM through demanding its provision in proposals to Cabinet. However, at Stage 2 there was a need to shift the focus from adopting new procedures to promoting evidence use and quality. Through establishing a broad approach to policy formulation that included EIPM but only as a part of a broader whole, this created a situation where procedures were being applied but the evidence-focused aspects were not necessarily the main area of change.

There is limited evidence that BCURE in Pakistan will catalyse a step change in the use of evidence across government. Evidence from across BCURE suggests that mass training alone is insufficient to promote routine evidence use; and while the project also conducted policy pilots and other activities, these did not join up with the training in a significant way (discussed in the single ministry impact pathway). Although EIPM training has been institutionalised in national courses – a significant achievement – this is unlikely to catalyse broader changes in practice in the absence of top down incentives and support structures to facilitate trainees to apply their learning.

Spotlight 4. Establishing a sustainable national actor through 'learning-by-doing'

BCURE in Zimbabwe was delivered through local partner ZeipNET. The programme was set up as a consortium model, with the aim of establishing ZeipNET as a credible Zimbabwean institution that could continue to promote EIPM once the BCURE programme ended. The evaluation found strong evidence that ZeipNET's capability to deliver EIPM capacity support has been strengthened and its profile raised, through formal training alongside ad hoc technical support to enable ZeipNET staff to 'learn-by-doing' and develop new technical and logistical skills. Programme staff felt that the programme was successful in building ZeipNET's capabilities because of lead partner INASP's existing relationships with ZeipNET's two main staff members, and the fact that these individuals had significant previous experience in the field of EIPM, were passionate about the goal of improving evidence use in policymaking, and already had knowledge and contacts from previous roles (in government and research) to draw on. Externally facing activities proved crucial to building ZeipNET's profile: several organisations introduced to ZeipNET through knowledge cafes or policy dialogues subsequently made requests for support or collaboration. However, a major concern to ZeipNET's ongoing sustainability is the resource-constrained environment of Zimbabwe. Several organisations and government ministries had expressed interest in partnering with ZeipNET to deliver EIPM training and other activities, but none of these organisations had resources to fund training, and there was some concern that donor resources are drying up.

In Kenya, AFIDEP's leadership of the SECURE programme has also built its credibility and capacity as a national EIPM actor – although this was not an explicit intended outcome and there was no formal capacity support within the consortium arrangement. AFIDEP has moved from being relatively unknown to being an acknowledged technical expert of research use in policy, a competency that is new in the wider health and parliamentary sector in Kenya. The relationship between the Ministry of Health and AFIDEP has also continued through ongoing membership of health research technical working groups, and joint presentations at events and proposal development. As a result of AFIDEP's leadership of the BCURE programme, the health R&D Unit has invited staff to sit on the national health research technical working group. Sector stakeholders have also indicated that AFIDEP has helped the Kenyan health research community to better understand the need to translate research to support decision and policymaking. Establishing AFIDEP as a credible actor in this field has meant that BCURE has left a legacy of a trusted civil society partner, which can continue to promote EIPM from within government and across research-policy networks.

> *It really helps if there is an external organisation whose job it is to think about how to help government to catalyse the evidence process, as we will have challenges of capacity for the long-term. (Ministry of Health stakeholder, Kenya).*

7. Impact pathway 3: Support to parliament

Parliamentary settings poses a different set of issues and challenges – as parliaments do not make policy, but can play an important role in interrogating it and holding line ministries to account. BCURE worked to build capacity for evidence use in parliamentary settings in Kenya and Zimbabwe. This section discusses their experiences, and lessons for future programmes.

7.1 What were the drivers, opportunities and risks for EIPM in the Parliaments of Kenya and Zimbabwe?

In both countries, Parliament had a clear mandate to scrutinise legislation and policies, substantial ongoing capacity support programmes, and well-established research services. Zimbabwe's parliamentary strategic plan aims to improve human capital to analyse and produce evidence for input into legislation, including through a strong research unit and committees. Substantial donor resources were directed at capacity building within Parliament, including a multi-donor Parliamentary Support Programme. These established routes for donor engagement provided a clear entry point for VakaYiko, who signed a 5-year MoU with Parliament at the start of the programme (however, the plethora of other capacity support also served to reduce the scope for a small programme like VakaYiko to make an impact). Recent and ongoing initiatives to strengthen Parliament include establishing a full time M&E Unit, holding public Open Days, maintaining ISO quality certification and joining the African Parliamentarian's Network on Development Evaluation. All of this activity suggests existing high-level buy-in for EIPM within Parliament, providing an opportunity for VakaYiko to align with and build on ongoing processes of reform.

In Kenya, the 2010 Constitution provided Parliament with greater constitutional powers for scrutinising legislation and policies, and approving and monitoring expenditure, which significantly increased demand for reliable information, data and evidence. Parliament has an established research services unit, with an explicit mandate to provide impartial, accurate and robust information to parliamentarians, which provided an entry point for SECURE. Prior to 2015, the PRS department had ten staff, with just six analysts, but the expansion in Parliament's mandate prompted the department to increase to 33 staff. Research unit managers were already proactively engaged in improving the use of evidence in research products. Researchers from the unit are active members in the Secretariat for each committee, alongside the clerks and legal counsel. The Kenyan Parliament is also well-linked into networks of African parliamentarians, and has an active training and capacity development programme supported by the Centre of Parliamentary Studies and Training, part of the Parliamentary Services Commission. Finally, greater public communication and transparency from Parliament means that citizens are better informed about their MPs' behaviours and voting records, and committees' proceedings, and are demanding improved performance, likely to be reinforced by new educational criteria and codes of conduct for Kenyan MPs joining after 2017.

However, in both countries, parliamentary powers (and incentives) to hold the Executive to account are limited by political influence, and active engagement with evidence is constrained by MP capacity and resource challenges. In Zimbabwe, the partisan nature of politics affects both the demand for evidence and the quality of parliamentary debate, while Portfolio Committees do not always have 'teeth' to hold ministries to account. Parliament is often perceived as 'rubber stamping' policy rather than holding the Executive to account. The capacity of MPs to engage with evidence can be variable, with challenges in building and maintaining capacity in time to make a difference before the next electoral cycle. The research department and other areas of Parliament are under-

staffed and face various constraints due to severe resource limitations in a broader context of economic stagnation.

In Kenya, political influence is also a major influence on decision making, exerted across all the branches of the state through patronage and regional alliances, which can lead to political policymaking, rather than decision making in support of government effectiveness and accountability to citizens. This constrains the extent to which evidence use can balance political decision making and enable robust scrutiny of the Executive. Weak procurement systems also mean that corruption remains a major barrier to effective public life. As in Zimbabwe, resource constraints and variable capacities of MPs to engage with evidence create further challenges.

Box 18. Summary of BCURE support to parliaments

In **Zimbabwe**, Parliament consists of the Senate (with 80 MPs) and the National Assembly (with 270 MPs). The role of Parliament is to deliberate on and pass laws, scrutinise government performance and to hold the Executive to account for the manner in which public policies and programmes are managed. Oversight of public resource allocations and policymaking in Parliament is delegated to Portfolio Committees, which consider bills and statutory instruments, and monitor and investigate policies and budgets relating to governmental departments. Portfolio Committees, supported by the research department, have a clear mandate to hold ministries to account, and consider evidence through investigatory processes. The Constitution of Zimbabwe emphasises that all agencies of the state and government are accountable to Parliament, and provides powers to summon ministers to Portfolio Committees to answer questions.

BCURE trained 20 staff members in Parliament, including the whole research department. It also ran a Learning and Exchange programme involving residential visits to the Parliaments of Ghana and Uganda. Technical support was also provided to help staff put Action Plans (developed during the EIPM training) into practice. These were intended to contribute to stronger organisational systems and processes to access evidence and engage in dialogue, in order to further promote individual behaviour change. Mentoring activities iteratively developed and changed over time in response to changes in priorities as well as logistical challenges.

In **Kenya**, Parliament consists of the National Assembly (with 350 members) and Senate (68 members), independent offices and commissions, and an independent judiciary. Both Houses of Parliament have a proactive role in law-making by scrutinising bills and policies presented by ministries (the Executive), and raising Parliament's own bills through members and departmental committees. In Kenya's devolved system of government, the National Assembly plays a key role in scrutinising, approving and monitoring expenditure by the national and county administrations through sectoral committees. The Senate scrutinises the work of the assembly, approving all bills affecting devolved functions such as health. In this way, both legislative houses check the power of the other, and of the Executive. The President can decline to promulgate a bill into law, but can, in theory, be held to account by an independent judiciary. The Parliamentary Research and Analysis Unit (PRS) supports both houses, providing over 50 committees with technical analysis of policies and budgets; drafting policies and bills, as well as supporting a range of committee-led enquiries.

BCURE trained 11 parliamentary researchers, through a five-day residential course, and supported the researchers through follow-up mentoring to help them complete policy briefs. The programme also involved learning and exchange visits, including an internship for two trainees to the UK Parliamentary Office of Science and Technology (UK POST).

7.2 What happened and why?

Figure 6 depicts how BCURE worked towards a step change in evidence use through providing support to parliaments. It does not represent what any individual project did, but rather synthesises evidence from across BCURE and the broader literature on how and why capacity support can lead to change. The diagram summarises what the evaluation has learned about how capacity support can contribute to EIPM in a parliamentary setting, but it is not a fully tested theory, as projects made different degrees of progress towards the intended impact. It is therefore intended as a broad 'road map' for future programmes working to promote EIPM in parliaments, rather than a definitive prescription.

Below the diagram, Table 7 presents an overview of how far each of the outcomes were achieved by the BCURE projects (EQ 1) and BCURE's contribution (EQ 2). These findings are unpacked throughout the rest of this section, which also explores how and why BCURE did or did not make a difference (EQ 3).

Figure 6. 'Support to Parliament' impact pathway

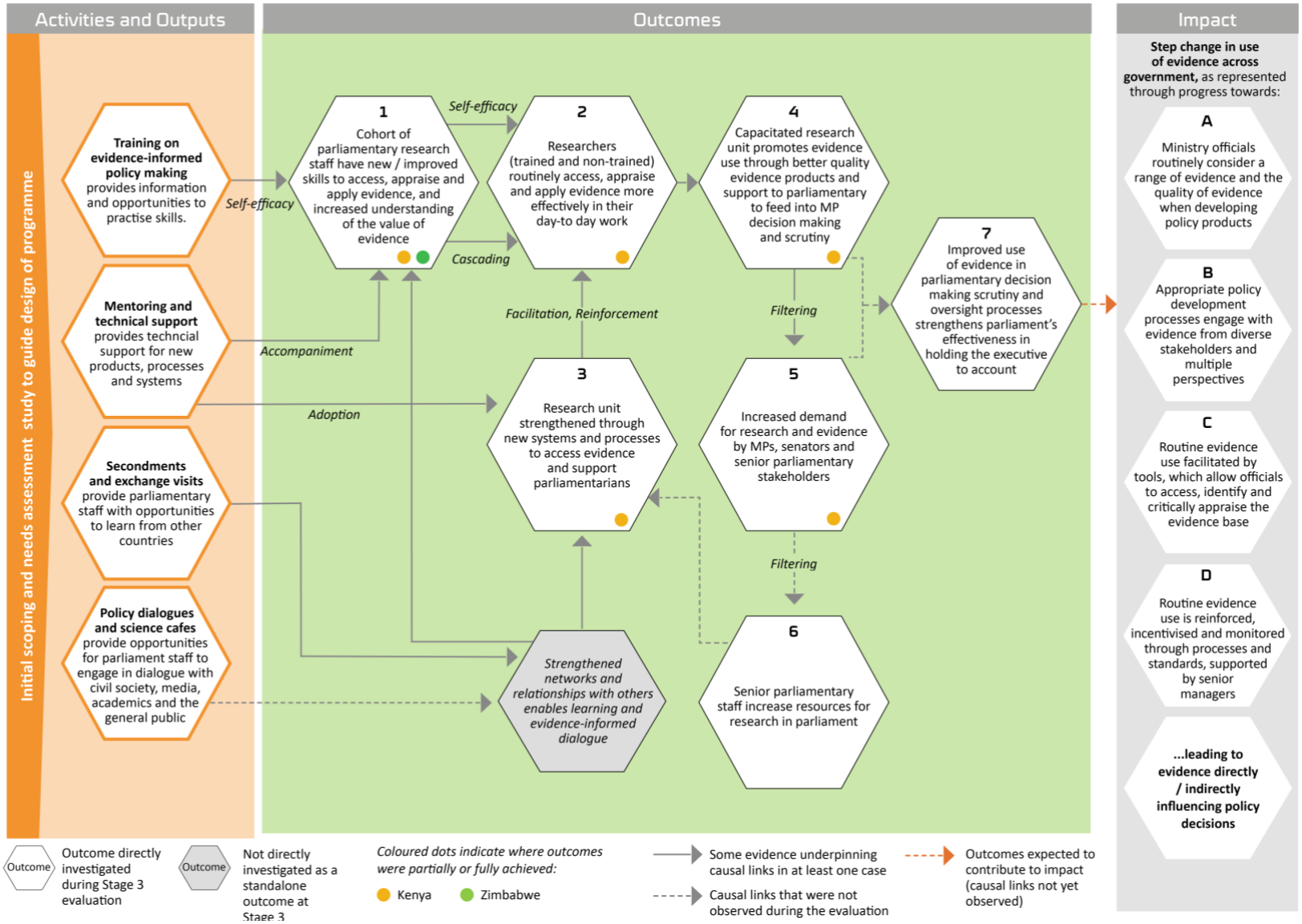


Table 7. Summary of evidence for 'support to Parliament' impact pathway

Outcome	Summary of evidence for outcome (EQ 1) and BCURE contribution (EQ 2)
1. Cohort of parliamentary research staff have new/improved skills to access, appraise and apply evidence, and increased understanding of the value of evidence	Strong evidence from **Kenya** that trained analysts had acquired new skills and that BCURE had made a crucial contribution to this. Some evidence from **Zimbabwe** that the programme had contributed to an increase in specific research-related knowledge and skills, alongside other, larger-scale capacity building programmes within Parliament.
2. Researchers (trained and non-trained) routinely access, appraise and apply evidence more effectively in their day-to day work	Strong evidence from **Kenya** that skills developed through BCURE training and mentoring were still being applied after the end of the programme. Strong evidence that staff not trained by BCURE also improved their skills in evidence use, and that BCURE had made an important contribution to this through supporting an internship to the UK Parliament which had led to interns introducing new evidence products and processes within the research department. Limited evidence that BCURE had contributed to routine shifts in evidence use in **Zimbabwe**, where examples of change seemed largely to do with non-BCURE capacity support. There is also no evidence from Zimbabwe of change in the behaviour of non-trained staff.
3. Research unit strengthened through new systems and processes to access evidence and support parliamentarians	Strong evidence from **Kenya** that the research unit was strengthened through new organisational initiatives developed by BCURE interns to UK POST. However, although the EIPM guidelines developed by the programme were officially adopted and approved by Parliament, they were not being used, and an EIPM curriculum had not been adopted by the parliamentary training agency. No evidence from **Zimbabwe** that the organisational processes developed by BCURE had strengthened the research unit, as they had not been adopted by Parliament.
4. Capacitated research unit promotes evidence use through better quality evidence products and support to parliamentary to feed into MP decision making and scrutiny	Some evidence from **Kenya** that BCURE's support to the research unit led to improvements in the quality, efficiency and take-up of their services. Some evidence from **Zimbabwe** that the capacity of the research department had improved in recent years – but there is no evidence that the programme made a significant contribution to this as most examples of change related to other, larger donor capacity support programmes.
5. Increased demand for research and evidence by MPs, senators and senior parliamentary stakeholders	Some evidence from **Kenya** that improved research services are stimulating a growing recognition of the value of evidence in Parliament. While BCURE made some contribution to this through improving the quality of work and thus increasing the profile of the research unit, higher-level drivers of evidence use are also giving momentum to the evidence agenda in Parliament. Views were mixed in **Zimbabwe** on how far MPs were requesting evidence from the research unit, but there is no evidence that the programme influenced any increases in demand.
6. Senior parliamentary staff increase resources for research in Parliament	No evidence from **Kenya** or **Zimbabwe** that BCURE has catalysed increased resources for research in Parliament.
7. Improved use of evidence in parliamentary decision-making scrutiny and oversight processes strengthens Parliament's effectiveness in holding the Executive to account	No evidence from **Zimbabwe** that BCURE has made progress towards improving parliamentary scrutiny and oversight. In **Kenya**, BCURE made an important contribution to improved use of evidence in the research department, although this has yet to become fully routine and there is no evidence that this has in turn contributed to improved use of evidence in decision making across Parliament.
Impact: Step change in use of evidence across government line ministries	Limited evidence of progress towards impact level change in either **Zimbabwe** or **Kenya**. In both countries, there are questions around how far Parliament is able to meaningfully hold the Executive to account given the variable ability and motivation of MPs to engage with evidence, and the fact that decision making by MPs remains a highly politicised process.

In both Kenya and Zimbabwe, EIPM training and learning exchanges contributed to a cohort of parliamentary research staff developing new or improved skills to use evidence (**O1**). In Kenya, this led to researchers (trained and non-trained) routinely accessing, appraising and applying evidence in their work, while in Zimbabwe, only a few researchers had acquired new skills or felt able to put them into practice (**O2**).

In Kenya, there was strong evidence that BCURE's support through training and mentoring had been effective in building research skills that were still being applied after the end of the programme. The response to the EIPM training was considerably more lukewarm in the Zimbabwean Parliament. In Zimbabwe, although there was some evidence that researchers had gained new knowledge and skills (including how to search for information, reference appropriately and write reports), only a few researchers felt they had been able to put these new skills into practice. In contrast, BCURE was able to make a substantive contribution to evidence use in the Kenyan PRS because it came at the right time when the unit was expanding, in a context where evidence use was already recognised by senior managers as being core to their mission to provide accurate, impartial advice to Parliament. The expansion of the PRS team provided an opportunity for BCURE to quickly establish a relatively large cluster of EIPM trained officers 'from scratch' (about a third of the total staff), who were already highly skilled and educated in other professional fields. There was some evidence that non-trainees were also improving their skills in evidence use, influenced in part by seeing colleagues' improved evidence products, and then reinforced by in-house training on EIPM and new processes developed by the UK POST interns, including more structured peer review processes, policy brief templates, and quality assurance guidelines.

> *Whenever we have in-house training, that issue of evidence use has always featured into our trainings. As a research department this is something that we have really embraced. (Programme participant, Parliament, Kenya)*

In Kenya, the skills offered by BCURE were seized upon because they were felt to bring something practical and directly relevant to help new recruits rapidly get to grips with a demanding job – in contrast with Zimbabwe, where the research unit was already established, and where the training was seen as insufficiently tailored to trainees' needs. In Kenya, a technical focus on evidence use throughout the policy cycle had not been offered through previous training in Parliament; whereas several trainees in Zimbabwe felt the training did not offer anything new. Overall in Zimbabwe, the sense was that the BCURE training course was a small programme that had made only a limited contribution to researchers' capacity, in a context where many larger-scale donor-supported capacity building opportunities are available. In addition, the training was seen as much less relevant to other staff outside of the research department, for example clerks and IT staff.

Box 19. Cost-effectiveness of learning exchanges and secondments

Three BCURE learning exchanges were considered within the evaluation. The Kenya and Zimbabwe projects both ran parliamentary learning exchanges; while in Bangladesh BCURE facilitated an exchange visit to Indonesia for high level government staff. The cost of learning exchanges and secondments varied significantly. In the case of Bangladesh, the cost per learning exchange participant was £1,731, while in Zimbabwe it was £6,000. In Kenya the cost per secondee to the UK Parliamentary Office of Science and Technology (UK POST) was £6,500.

The Stage 2 evaluation found that learning exchanges do not always have instrumental value, but rather tend to shape participants' thinking and provide 'experiential learning'. Therefore, while of value, they are relatively low impact activities when considered as standalone interventions. In light of this, from a cost-effectiveness perspective, the implication is that the costs should be kept relatively low and the exchanges well-synergised with other activities. While Bangladesh seem to have pitched the costs right, the Zimbabwe costs are on the high end.

In the case of the Kenya secondments to UK POST, while the cost per participant is high, the two researchers that participated worked at UK POST for a month and received mentoring on their return to implement organisational initiatives to improve evidence use across the PRS unit in Parliament. The evaluation found that the secondment had resulted in transformational change for the participants and contributed to them putting into effect organisational changes. From this perspective, while the costs are higher, the value that has been generated has been significant.

In both countries, BCURE provided mentoring and technical support to a range of organisational initiatives to promote and embed evidence use, with only partial success (**O3**).

In Kenya, there is strong evidence that organisational initiatives were successfully developed through internships to the UK Parliament, which helped to facilitate evidence use and cascade evidence skills to non-trained colleagues; while in Zimbabwe some organisational initiatives were successfully trialled, but ultimately were not adopted by Parliament. In Zimbabwe, new initiatives included an Evidence Roundtable seminar series and support to help researchers access e-resources. However, there was limited evidence in either case that the initiatives had been adopted by Parliament. This was in part because technical support was relatively ad hoc and built around standalone activities and products – this component was not sufficiently resourced to provide comprehensive ongoing and flexible support, and ensure that activities built on and reinforced each other. It is also unclear whether Parliament had the resources to adopt more ambitious initiatives like the Evidence Series in the longer term, even if BCURE had facilitated further events, in the deeply resource-constrained environment of Zimbabwe where a large number of parliamentary initiatives are donor-funded. Communication between the programme and the research department also appears to have been an issue, feeding into a lack of ownership which may have further limited the success of the mentoring programme.

In contrast, in the Kenyan PRS, the UK POST interns were able to introduce new evidence initiatives to the PRS unit because it had an established culture of quality improvement – meaning that the interns had the managerial support and organisational opportunities they needed to develop procedures and processes to facilitate and reinforce evidence use. The team-working culture in PRS further encouraged BCURE trainees to share skills with non-trained colleagues.

Among the reasons for more visible success in Kenya was the ownership of evidence initiatives by PRS managers and more sustained technical support by the programme. In Kenya, BCURE's technical support was more comprehensive and joined up across the PRS unit, with support provided to produce evidence guidelines for Parliament, as well as ongoing mentoring to individual

trainees and providing further skills support through the in-house workshops. The evidence guidelines were produced through a collaborative process, with the involvement of senior managers, and led to a good quality and user-friendly draft document. As a result of strong management ownership, the guidelines were approved as official parliamentary procedure by the Parliamentary Commission for use by all MPs and clerks. Given the bureaucratic and procedural complexities of the parliamentary system, this was a considerable achievement – although by the end of the programme, the guidelines had not yet been printed and released for formal use. A key factor was the proactive advocacy of PRS managers at board level. The good level of institutional ownership of the guidelines can be directly attributed to BCURE's collaborative and flexible technical support, which was valued by stakeholders. BCURE's ongoing support played a more successful 'accompanying' role to PRS, helping to embed reforms to embed evidence use. Peer review and the template had been implemented, with ongoing support from BCURE, and there was evidence that this was facilitating and reinforcing evidence use, in a context where managers were providing strong encouragement to improve quality and efficiency, given the volume of research and evidence products that PRS needs to produce for both Houses of Parliament.

Box 20. Summary of what worked for who and why in providing organisational support for the adoption of new tools and systems (EQ 3)

In Zimbabwe, Parliament has not adopted new tools or systems introduced by the programme (M, O3) in part because BCURE did not sufficiently embed them as a result of resource limitations (I). In Kenya, Parliament has adopted (M) the tools and these have strengthened the research department (O3), because BCURE support was ongoing, flexible, and joined up individual and organisational support, meaning that BCURE was able to accompany (M) the unit to embed the approaches. This was enabled by the committed and hands-on advocacy and encouragement by senior managers (C), who could see quality improvements from the use of tools (O4), as they facilitated and reinforced (M) the use of evidence by researchers. In Zimbabwe, given resource constraints in Parliament, it is unclear whether Parliament has the resources to adopt an initiative like the Evidence Series in the longer term even if ZeipNET had facilitated further events (C). As a result, there has been no opportunity for new tools and systems to facilitate use of evidence in Parliament (M) or reinforce evidence use through positive or negative incentives (M).

In both countries, there is evidence that the capacity of parliamentary research units has improved in recent years – although in Zimbabwe there is limited evidence of programme contribution (**O4**). In Kenya but not Zimbabwe, this has contributed to increased demand for evidence from parliamentarians (**O5**), but in neither case have there been any increase in resources for research (**O6**).

In both Kenya and Zimbabwe, there are signs that research departments are more effectively working with evidence – in Kenya there is some evidence that BCURE contributed to this, but in Zimbabwe there is limited evidence that the programme made a significant contribution in a context of many larger donor capacity support programmes. In Kenya, BCURE's support to skills and systems in PRS has led to improvements in the quality, efficiency and take-up of their services, increasing the unit's profile in the parliamentary system. Most respondents felt that there had been a significant improvement in the quality of their products and outputs, noted by MPs. Both researchers and senior managers also noted improvements in efficiency. Trainees' use of evidence was further reinforced by career incentives made possible by the parliamentary context, where senior management recognition of good performance leads to assignments in high-profile committees, including international trips, which in turn bring salary enhancements in the form of expenses. The evaluators noted that staff seemed motivated and enthusiastic, and the unit has

sufficient numbers of staff to be able to assign a researcher to the 50 or so active committees across both Houses of Parliament. However, despite some tangible outcomes in terms of behaviour change and quality improvements, these were still seen as early steps in the department's use of evidence.

In Zimbabwe, there are similar signs that the skills and systems of the research department have improved in recent years, but there are many other donor programmes also providing capacity support, larger than BCURE, so evidence of BCURE's contribution was limited. There are also still serious capacity constraints within the unit, linked to understaffing, with senior staff reporting that the team is extremely stretched, trying to serve multiple committees and responding to ad hoc requests of MPs. Time to do independent research is very limited and researchers frequently cannot develop sectoral expertise because they are spread too thin and are sometimes working on topics they are unfamiliar with. In this context, the BCURE support contributed in a limited way to enhanced individual skills but did not make headway with organisational systems in the face of staff and resource constraints.

Despite growing appreciation for the work of the research departments in both countries' parliaments, there are no signs that improved products are influencing an increase in demand for evidence, and therefore increased investment in research, even in a context where there are high-level drivers in favour of increased use of evidence. In Zimbabwe, although some parliamentarians expressed appreciation for the research unit, there was limited evidence that BCURE's support has led to improved products that in turn stimulated greater demand, or that this is likely to lead to greater resources being assigned to research in Parliament. The work of the research unit is viewed as important in Parliament, and its contributions are valued, including producing research papers and providing inputs to portfolio committee reports, particularly the Constituency Profiles. However, given that BCURE only made a limited contribution to individual skills, any increase in demand for research is more likely to be a result of institutional drivers of EIPM. MPs' interest in evidence may also be in part driven by the desire to avoid being refuted in the media, as public access to information (e.g. through social media) improves.

In Kenya, there were indications that an improved service from PRS is stimulating a growing recognition of the value of evidence in the wider Parliament, although higher-level drivers of evidence use are also giving momentum to the evidence agenda. Greater transparency around Parliament's work, including a focus on MPs' performances in debates, is creating incentives for MPs to be better informed, factors that contributed to the formation of the Evidence Caucus in 2015 (an informal grouping of MPs who are interested in promoting evidence use among members). The idea for the caucus preceded BCURE, but enabled lead partner AFIDEP to take the opportunity to create a second evidence project in Parliament, resourced by a different funder. Establishing the Caucus Secretariat in PRS has provided an institutional focal point for the evidence agenda in Parliament and has also allowed AFIDEP to continue to support EIPM in PRS by contributing to in-house workshops funded by Parliament. BCURE's support, alongside these factors, does seem to have made a contribution to increasing the profile of PRS – for example, MPs are using PRS more, and the Speaker has requested that researchers now also attend second readings of bills in the chamber. However, the potential for further investment in the PRS unit, e.g. to upgrade the IT infrastructure to publish committee reports, seems limited. Within the wider parliamentary system, PRS is still a very small player and does not have its own budget as a small unit within a much larger Joint Services department. Furthermore, BCURE has not succeeded in getting the EIPM curriculum taken up by the parliamentary training agency, Centre for Parliamentary Studies and Training, which was the main strategy for continuing EIPM support after the programme.

Box 21. Summary of what worked for who and why in influencing parliamentary stakeholders to demand evidence more, and invest in research (EQ 3)

While several stakeholders in both Zimbabwe and Kenya felt that producing good quality work can 'filter up' to impress MPs (M), which in turn demonstrates the value of evidence and leads to increased demand, there is limited evidence that this is actually happening in practice (O5). Filtering up is inhibited by capacity constraints of MPs (C) and the fact that the research units have limited power to influence parliamentary processes such as fact finding missions in committees (C). In Zimbabwe, understaffing within the research department makes it difficult to respond to demands (C) and reduces the potential for 'filtering up (M), while in Kenya, there is sufficient capacity to serve all the parliamentary committees and produce a high-volume of work (C) which makes filtering up of good products (M) more like to happen. However, in both countries, filtering up is also likely to be inhibited by the broader political context – in which issues in Parliament often get debated along political lines and MPs have limited incentives to challenge the government and are commonly seen as 'rubber stamping' government policy (C). Without support to strengthen MPs' ability and incentives to demand evidence (I, C), improving quality of parliamentary research support was not sufficient on its own to catalyse further investment in research in either country (O6).

Although several stakeholders in Kenya and Zimbabwe feel that evidence is becoming more important within parliaments, there is no evidence to link this to BCURE support. Overall there is limited evidence that BCURE has contributed as yet to improved use of evidence in parliamentary decision making and oversight (**O7**), leading in turn to a step change in use of evidence across government line ministries (**impact**).

In **Zimbabwe**, there is very little evidence to suggest that the programme has contributed to a significant shift in the quality of work produced by the research department, or made progress towards its anticipated longer-term outcomes of improving parliamentary scrutiny and oversight processes in order to contribute to more routine, embedded and transparent use of evidence across the Government of Zimbabwe. Considering its short duration and resource limitations, this ambition was beyond the realistic scope of the programme. However, in a context where there are many existing drivers of EIPM including ongoing reforms and high-level champions, as well as substantial donor resources directed at capacity building across the institution, BCURE has helped to better equip the research unit to respond should demand for its services increase in the future, although staffing and resource constraints could still seriously limit their contributions.

In **Kenya**, BCURE made more significant contributions to improved use of evidence in PRS – although this has yet to become fully routine, which suggests the need for longer-term ongoing support. In both countries, there are also deeper-seated questions around how far parliaments are able to meaningfully hold the Executive to account given the variable ability and motivation of MPs to engage with evidence, and the fact that decision making by MPs remains a highly politicised process.

8. Conclusions

This report has presented the summative findings from the independent evaluation of the Building Capacity to Use Research Evidence (BCURE) programme. Below we draw out conclusions and six overarching lessons on what works, for whom and why to build capacity for evidence-informed policymaking.

BCURE was only four years long; a very short time to observe change in government behaviour and processes. BCURE's scope was broad, its goals were ambitious, its time frame was short and its resources were restricted – when dispersed across 12 countries and when compared to other governance reform programmes. Some activities were too small scale and ad hoc to catalyse change within large ministries or parliaments, or too disconnected to combine to promote change at a national level. As a result, although the evaluation has observed good progress across a number of projects, it is too early to judge the extent to which BCURE has contributed to a step change in the use of evidence. The evaluation noted three levels of success across the BCURE portfolio:

- **Significant progress towards catalysing change at scale: Bangladesh, Kenya (Parliament), and Sierra Leone.** These projects involved 'top down' activities to establish procedures and incentives for evidence use at an organisational level, combined with 'bottom up' capacity building for technical staff – a model that appears to have significant potential to catalyse long-term progress towards EIPM. However, the evidence is tentative across all of these settings, especially in Sierra Leone which was not included in the Stage 3 evaluation. Although there is strong evidence of senior-level ownership, it is too early to tell whether promising early progress will continue. This depends on continued political leadership, high-level incentives and resource mobilisation, which are all potentially fragile and subject to change.
- **Pockets of success around specific policy processes and capacitated units: Kenya (Health), Zimbabwe (Youth), Pakistan policy pilots, South Africa evidence map.** Across most BCURE projects, there are examples of improved capacity at an organisational level, or good quality policy pilots or tools that have showcased the value of evidence use and led to small-scale adoption. This happened where BCURE projects identified clear windows of opportunity and provided collaborative support within settings where there were existing organisational incentives for change. However, these examples did not catalyse incentives for evidence use at a system level, which may impede their long-term influence.
- **Ad hoc and 'one dimensional' change: all six projects.** Across the BCURE portfolio, there are many examples of individuals applying new knowledge and skills within specific policy processes – but while these are important demonstrations of individual behaviour change, they are ad hoc. In Pakistan, the institutionalisation of EIPM in national civil service training is a significant achievement, but because it was not joined up to other activities or reforms, it has limited chance of contributing to a step change on its own.

The report began by identifying three overarching ways of working that underpinned success across the BCURE programme – thinking and working politically, building capacity at multiple levels in the system, and accompanying change not imposing it. Through exploring what worked for whom and why in the diverse BCURE contexts, the report has demonstrated why these ways of working are important, and how they help to catalyse key mechanisms (see Box 22) that lead to change. Programmes that aim to build capacity for evidence use are often designed around specific activities, such as training or technical support. Our findings suggest the need to think beyond activities, and instead begin by considering the key ways of working and mechanisms that underpin successful capacity development – and then think through how best to catalyse these in a particular context. The six lessons below, and corresponding recommendations in Section 9, unpack these insights.

Box 22. Mechanisms underpinning success in BCURE

The report identified six key mechanisms that, when catalysed, led to positive changes around the use of evidence, although not all of them were present in any one project. As the impact pathway analysis illustrated, the key mechanisms do not operate in insolation, but instead work together to catalyse change, and build on each other so that where one mechanism operates it often creates a conductive context for another mechanism to 'fire'. These mechanisms are derived from well-established theories from psychology, sociology, development studies and governance – referenced in 'insights from the literature' boxes throughout the report. In this section, we present the evaluation's revised and tested theories about how these mechanisms (M) are generated in particular contexts (C) and through particular features of interventions (I) to lead to capacity outcomes (O).

Accompaniment: where an external partner provides tailored, flexible and responsive support to a government institution through a process of reform, characterised by a high-level of trust, as opposed to a more traditional supplier/consumer model where ad hoc support is provided through one-off interventions. This often involves co-producing tools, systems or policy products.

Self-efficacy: where providing information, opportunities to practise skills, coaching or technical support builds individuals' confidence in their ability to do their jobs or achieve a particular goal. This is akin to feeling of 'now I know how... (to find the evidence I need, to weigh up sources, to communicate evidence effectively).'

Facilitation: where a tool, system or process for EIPM facilitates government officials to do their jobs or undertake a task more easily or efficiently.

Reinforcement: where rewards or other forms of control create incentives that motivate officials to work in a particular way. Positive reinforcement includes rewards and encouragement, while negative reinforcement includes reminders, audits and mandatory requirements.

Showcasing: where providing good examples of evidence tools or processes demonstrates the value of an evidence-informed approach, which leads to them being adopted elsewhere.

Adoption: where senior government stakeholders decide to adopt a new tool, system or process for EIPM to help standardise EIPM within a government institution. This can be on a small scale (a unit deciding to adopt a new template to standardise policy briefs) or a large scale (a government deciding to adopt a revised procedure for policymaking across all its line ministries that requires engagement with evidence). Adoption can happen for many reasons, and there is a risk of 'isomorphic mimicry' – where a new tool or system is adopted on the surface in order to access donor resources, without actually changing day-to-day practice.

The evaluation also identified a further mechanism that implicitly underpinned several BCURE projects, but which has not (yet) catalysed in practice on a large scale:

Critical mass: where changes in practice among a sufficient number of government officials diffuse out to influence colleagues' behaviour, and the rate of adoption of new behaviours becomes self-sustaining. This diffusion may happen through **cascading**, where government officials formally cascade their new knowledge on EIPM through introducing new ways of working or new structures and processes. Or it may be through **filtering out** or **filtering up:** where improvements in evidence use by government officials leads to recognition of the value of an evidence-informed approach among colleagues (filtering out) or senior management (filtering up) which in turn influences' colleagues behaviour, or increases senior-level support for evidence-informed ways of working and/or organisational reforms to promote EIPM.

Lesson 1: BCURE highlights the importance of thinking and working politically, taking a political economy lens to assess the context and the potential for change. Choice of an entry point to a government setting might be opportunistic, but success depends on a thorough analysis of the incentives and disincentives to consider evidence in the context.

All six BCURE projects were superficially a good fit with government-owned agendas around EIPM, had evidence of demand from senior leaders, and were tailored to align with ministry needs and requirements through needs assessments. However, in contexts where government resources barely cover salaries and there is a reliance on donor funds for implementation, government stakeholders will often be receptive to programmes that bring resources, especially as civil servants are often explicitly mandated to mobilise donor resources. As BCURE was seeking to catalyse and institutionalise genuine reforms to promote evidence use, scoping activities should have looked beyond 'face value' statements of interest, and considered deeper internal political economy dynamics within ministries, which shaped the potential for catalysing change.

As BCURE progressed, implementers became more alert to political economy dynamics – better understanding how they could align with existing incentives to give EIPM skills, capacities and systems the best chance of being genuinely embedded in government systems. 'Thinking and working politically' is an idea that has gained considerable traction in the international development field in recent years, emphasising the importance of identifying political 'windows of opportunity' where incentives align to create genuine interest in reform and give it a chance to take root. The BCURE evaluation findings fully align with this school of thought and underscore the importance of taking a political economy lens to optimise the effectiveness of interventions to support the use of evidence in government settings. The success of this approach is particularly highlighted in relation to policy pilots in the single ministry pathway, where partners in Kenya, Bangladesh, South Africa and Pakistan all succeeded in identifying windows of opportunity around key policy areas or political challenges, capitalising on existing work and partnerships, identifying allies, and leveraging external resources.

Thinking and working politically requires a deeper consideration of how gender and social inequalities might constrain individuals' power to influence change. As discussed in Spotlight 1, most BCURE projects missed the opportunity to integrate gender perspectives into their capacity support, including seeking opportunities to provide tailored support to build women's leadership and influence in contexts where women officials are likely to experience more difficulties in influencing change than their male counterparts. Collecting gender-disaggregated evidence of programme outcomes is an important aspect of this.

The evaluation does not make a judgement on whether programmes should work within a single ministry (impact pathway 1), seek to promote cross-governmental change (impact pathway 2) or work with parliaments (impact pathway 3). There are examples of more and less success across all three pathways, and there is insufficient evidence on the longer-term results of the different routes to impact given BCURE's short duration. Rather, the findings suggest the importance of choosing an approach based on an assessment of where political dynamics establish the potential to catalyse change, alongside existing relationships and networks that can give an external partner a 'way in.'

Lesson 2: Capacity support interventions should seek to accompany change, not impose it. Success depends on 'accompaniment', which requires a politically informed approach to help build strong local ownership, supported by a flexible approach to programme design, delivery and management.

The BCURE projects had most success where they 'accompanied' government partners in a flexible, tailored, collaborative way that promoted ownership, and strengthened partner capacity through 'learning-by-doing'. Accompaniment is critical to working politically, as it implies a responsive and evolving process of support, which is flexible enough to adapt to evolving challenges and opportunities in complex government contexts.

Different BCURE projects adopted this mode of working to varying degrees and at different scales. Where BCURE projects accompanied ministries or units through specific policy processes, this led to co-produced new tools or policy products which served to showcase the value of an evidence-informed approach, as well as supporting government partners to 'learn-by-doing'. On a larger scale, some projects accompanied a specific unit at ministry or cabinet level, promoting EIPM reforms through a collaborative model characterised by high levels of government ownership, spending significant time on building and maintaining senior-level relationships and buy-in, and responding to windows of opportunity where they arose. Where this approach was *not* taken, or was attempted but did not succeed, BCURE was less successful in catalysing ownership and getting reforms to take root in government systems.

Wherever BCURE was successful in accompanying reform, certain factors were key. In responding to high-level incentives and opportunities to support EIPM, BCURE projects formed partnerships with government departments or units who already had a mandate, and some authority, to promote evidence use, e.g. Cabinet Division, research units or M&E units. In some cases, partners were invited to accompany policy processes as a result of the relationships and trust they had built up through previous activities e.g. previous evidence-related programmes, research studies or technical working groups on priority policy issues. Individual champions within gateway institutions were often crucial – the three projects that worked across government all relied on a high-level, passionate and enthusiastic champion who acted as both a gatekeeper for the programme, and an advocate that paved the way for BCURE to embed reforms in government systems. Accompaniment is not straightforward, and projects are likely to face numerous blockages that need to be navigated, including staff rotations at technical and senior-level, corruption scandals, and changes in government priorities. In order for programmes to work in this way, there needs to be sufficient flexibility in the contracting model, to allow partners to respond nimbly to challenges and opportunities.

What does this tell us about what works for whom in what contexts?

CIMO 1. Where there is genuine interest in partnership from high-level government stakeholders, existing incentives for evidence use in policymaking, and a window of opportunity to catalyse reform (C), an external partner can accompany EIPM reforms (M) in a participatory and collaborative way, providing tailored, flexible and responsive ongoing support that evolves over time (I) in response to emerging challenges and opportunities (C). This mode of working is greatly helped by the presence of high-level, enthusiastic and committed champions (C), and can create a conducive context for the other EIPM mechanisms to operate through encouraging government ownership (O) and building trust in the partner to work alongside government (O).

Strong evidence in support of theory, from across all six settings

Lesson 3: Changing individuals' behaviour is the bedrock for EIPM, but requires more than building skills. It also requires establishing or harnessing incentives that reinforce changes in practice, working to build capacity at multiple levels of the system.

There was a genuine need to build technical skills in evidence access, appraisal and use across all the BCURE contexts. Individual capacity support mainly targeted the technical policy and research staff who are responsible for designing policy documents and developing research products that feed into policy formulation. However, BCURE's main success stories went beyond training or other individual-level capacity building, providing follow-up support to help to embed learning, promoting improved organisational procedures, co-producing tools or policies, or generating incentives to enhance use of evidence. The evaluation highlights the necessity of working with senior managers and government stakeholders – who may not be as involved in the technical side of evidence access, appraisal and use, but whose awareness and buy-in is essential to create an environment where technical staff are supported and incentivised to work in a different way.

Where BCURE led to more routine individual-level changes in evidence access, appraisal and use, this was often because projects succeeded in catalysing multiple mechanisms together: building staff self-efficacy, providing tools that facilitated people to do their jobs more easily, and tapping into or generating organisational incentives to reinforce behaviour change. Where training courses were directly relevant to participants' day-to-day work, training and follow-up interventions built self-efficacy through imparting new knowledge and skills, and raising trainees' confidence in their ability to perform their roles. In several countries, training and associated technical support also provided participants with practical evidence tools or processes that facilitated them to do their jobs more easily or more efficiently – such as the EIPM guidelines in Bangladesh and Cabinet templates in Sierra Leone, which provided a structured approach to policy formulation in contexts where this did not already exist. However, training and tools need to be closely targeted to officials who can use them, and directly relevant to their day-to-day work. In Pakistan and Bangladesh, where large cohorts of civil servants were trained, several participants said they were not working in roles where they could apply their learning. Finally, training seemed to be most effective in catalysing more routine (rather than ad hoc) behaviour change when evidence use was reinforced through organisational incentives such as supportive managers and senior staff, which motivated participants to apply their learning.

What does this tell us about what works for whom in what contexts?

CIMO 2. Where information is provided about the importance of EIPM and how to access, appraise and apply evidence, alongside opportunities to practise skills, this can generate self-efficacy (M) and lead to individual behaviour change (O). Behaviour change is more likely to be sustained where there are clear incentives that motivate participants to apply their learning and reinforce changes in practice (M) – this includes management support to encourage and provide space for participants to access, appraise and apply evidence, which in turn depends on political incentives that promote evidence use (C). Behaviour change is also more likely where activities are closely targeted to individuals who can apply their learning because it is directly relevant to their day-to-day work (I), and where activities are practical and participatory (I), provide practical tools (I) that facilitate trainees to do their jobs more easily (M), incorporate a focus on soft skills as well as technical skills (I), use knowledgeable, patient and confident facilitators (I), and tap into incentives to encourage participation (I).

Strong evidence in support of theory, from across all six settings

Lesson 4: Specific and targeted strategies are required for a 'critical mass' effect to catalyse. It is a common assumption that training a 'critical mass' of individuals will diffuse out to influence broader change – but interventions need to be designed and targeted in specific ways to leverage this effect.

Several BCURE projects were premised on an implicit theory that training cohorts of officials would catalyse a 'critical mass' effect that would influence shifts on a wider scale. However, even where BCURE was successful in establishing routine behaviour change, there is no evidence that this has coalesced to shift behaviour beyond the initial group of trained officials. As a result, there are only tentative lessons that can be inferred for future programmes – mainly relating to the factors that appear to block a critical mass from forming. If training is not directly relevant or there are missing incentives and organisational structures for EIPM, then individuals may be unable to apply their learning in the first place, as in Pakistan, and so there is little prospect of them influencing others. If individuals are too scattered across siloed units and divisions then this dilutes their opportunity to influence, as in Kenya. If trained officials are based in a unit that has limited power and resources, or if the programme works only with technical staff but not their managers or senior decision makers, this limits the possibility of influencing senior-level attitudes or behaviours, as in Zimbabwe. The only example of the 'critical mass' effect stems from Kenya, and is relatively small scale. Researchers were able to cascade new evidence initiatives within their own unit and from central to county level because senior managers supported and incentivised quality improvements, which motivated non-trained staff to engage with learning opportunities; and because the Ministry of Health was able to access resources to cascade training as this aligned with its mandate to build provincial level capacity.

Bangladesh is the most promising setting for a critical mass effect to emerge in future, as a high number of officials have been trained across government, and the project has established top-level incentives and reinforcement through Cabinet Division ownership. In this context, embedding the training course in national curricula to build EIPM skills on a large scale across the civil service may well catalyse more widespread change in future – but this is only likely if there continues to be high-level leadership providing incentives for trainees to apply their learning.

What does this tell us about what works for whom in what contexts?

CIMO 3. Where a cohort of officials start accessing, appraising and applying evidence more effectively, this can diffuse out to influence colleagues' behaviour (O) through a 'critical mass' effect (M). This is more likely when the cohort consists of a good number (I) of well-connected and clustered officials (C) in a unit with some reach and influence within the broader organisation (C), and where there are clear organisational incentives to use evidence (C) and senior management support to cascade learning (C) – potentially supported by a 'training of trainers' strategy (I)

Tentative theory, based largely on insights about blocking factors from Kenya and Zimbabwe, and insights from the wider literature

Lesson 5: Supporting practical examples of evidence tools or evidence-informed policy processes can showcase the value of evidence and catalyse 'learning-by-doing'. However, this requires identifying priority policy areas and problems, and establishing government ownership through an 'accompanied' process, rather than an external partner doing the work themselves.

Several BCURE partners worked to showcase the practical value of an evidence-informed approach through providing support to policy processes, or helping develop decision-support tools that enable officials to engage with evidence more easily. This proved one of the most successful interventions, leading to evidence-informed tools and policies in Kenya, Bangladesh, Pakistan and South Africa. This success was due to BCURE partners identifying a 'win-win' policy entry point, where there was a real need to solve a policy or service delivery problem, and where there was the potential to build on existing work and leverage external resources. It proved essential to secure senior support for the process, and involve stakeholders at the right level of seniority and with the right technical and interpersonal skills, from within and outside government. In most cases, BCURE provided flexible, responsive and tailored support to 'co-produce' policies and tools in partnership with government. This helped ensure ownership, which in turn made adoption of the resulting tool or process more likely – and it also supported individual-level capacity building through a process of 'learning-by-doing.' The policy pilot process offered an opportunity for strong synergies with other activities, including embedding skills gained through training, and using EIPM tools to provide a structure for the process that helped ensure evidence was considered in appropriate ways.

However, good quality policy products or useful data visualisation tools are only an early, although important, step towards evidence being used to inform decision making. In all the settings, the pilots were several steps away from this point. These subsequent steps were beyond the scope of BCURE to influence and were threatened by the various political and contextual barriers discussed throughout this report.

What does this tell us about what works for whom in what contexts?

CIMO 4. Where technical support is provided to incorporate evidence within a policy process, or develop a tool to improve evidence access, appraisal or use, this can generate high quality policies or products (O) that showcase the value of evidence for quality, performance and delivery (M) and lead to adoption (O) and diffusion (O) of the procedure or tool. This is more likely where external actors 'accompany' government partners to co-produce policies or tools in a flexible, responsive and collaborative way (I), where policies are high priority or tools address a recognised problem (C), and where tools are intuitive and interactive (I) and genuinely facilitate officials to make decisions and do their jobs better and more easily (M). However, adoption can be stymied by many factors including shifting political priorities or resource constraints (see CIMO 5).

Strong evidence in support of theory, from Kenya, Pakistan, South Africa and Bangladesh

Lesson 6: Promoting the adoption of system-wide EIPM approaches is important to lay the groundwork for future change. Combining capacity support at multiple levels, through accompanying change and tailoring support to the context, can help create the conditions for government partners to genuinely adopt tools and processes, and use these to develop their own EIPM initiatives into the longer term. However, this will only be sustained if the EIPM agenda continues to be aligned with wider political incentives.

An important long-term aim of any EIPM capacity building intervention is adoption: where government partners formalise new processes, tools or practices at an organisational level, which in turn go on to catalyse, deepen and incentivise individual-level change. However, there is a real risk that adoption will happen on paper, and not in practice. It is too early to say how genuine the adoption observed in BCURE will prove to be in the long term, given the risk of isomorphic mimicry in low and middle-income countries, especially in fragile contexts such as Sierra Leone. However, the evaluation illustrates that multiple mechanisms working together to create and reinforce change can create a context that helps make 'genuine' adoption more likely.

Adoption happened in BCURE on both a small and large scale:

- **Small-scale adoption** involved using and rolling out guidelines or EIPM support tools in specific units or sectors, as in South Africa and Pakistan. This happened where tools proved genuinely useful to officials' work, senior managers could see their value, and there was a clear institutional home for the tools going forward as well as resources for scale-up. Where EIPM tools have been genuinely adopted – in the sense that there are incentives in place and senior-level encouragement to use them – they can then facilitate and reinforce individual behaviour change, as the EIPM guidelines in Bangladesh have potential to do in future. Reinforcement does not come from a tool being adopted, but rather from it being actively promoted, and staff being supported to use it, as in Sierra Leone. Signing off guidelines as 'official' tools is not enough, as the example of guidelines in the Kenyan Ministry of Health showed.
- **Large-scale adoption** involved formalisation of a new tool or process on a government-wide scale, such as the EIPM guidelines in Bangladesh and the Cabinet procedures in Sierra Leone. EIPM training was adopted into national training institutes in Bangladesh and Pakistan through the support of high-level champions, in contexts with an established culture of civil service training. However, embedding a training course at a national level carries risks, as it will potentially move training away from the intervention factors found to catalyse individual-level behaviour change – in particular through diluting the courses, watering down the targeting, and removing any follow-up support. Without ongoing reform in the broader context that provides incentives for trainees to apply skills, institutionalising a training course is not enough on its own to ensure sustainability of reforms.

The deepest form of adoption is where capacity support catalyses further self-transformation – positioning a national unit to carry on promoting EIPM into the future. However, there is limited evidence that this has happened in BCURE. The example with the most potential for this is Bangladesh, where there appears to be genuine interest and ownership in Cabinet Division to continue promoting institutional reforms. However, it is too early to tell whether this will continue, especially in a context where frequent senior staff rotations can quickly erode buy-in. The experience of the DPME in South Africa suggests that 'genuine' adoption requires ongoing government ownership and resources to be catalysed beyond the project, as high-level incentives shift and new opportunities rise and fall in dynamic political environments.

What does this tell us about what works for whom in what contexts?

CIMO 5a. Where capacity support succeeds in showcasing the value of an evidence-informed approach, training course, tool or product (M) and/or generating tools that facilitate staff to do their jobs more easily (M), this can lead to a high-level decision to formally adopt the initiative to help standardise EIPM (O). Meaningful adoption is more likely where reforms have been co-produced by government and external partners through a flexible and collaborative process of accompaniment (C), and where there are high-level institutional and individual champions with a clear mandate for and interest in reform (C) who have access to resources to scale up or roll out the initiative (C). Adopted tools and processes, when attached to high-level incentives and encouraged through ongoing support rather than just a one-off directive (C), can then help reinforce (M) changes in practice at an individual and organisational level through both 'carrots' and 'sticks' (O).

Strong evidence in support of theory from Bangladesh, Sierra Leone, Kenya, South Africa and Pakistan. Insights on factors that blocked adoption in Zimbabwe also support this theory.

CIMO 5b. Where capacity support succeeds in catalysing high-level ownership and buy-in to EIPM (C), it can position an institution or unit to carry on promoting EIPM into the future (O), provided it is able to access resources (C) and buy-in is not eroded by high-level changes in priorities or staffing (C).

Tentative theory based largely on insights from the South Africa impact case study, and early observations in Bangladesh

As a final reflection, the evaluation team would like to acknowledge the invaluable opportunity we were given to follow the BCURE projects through their efforts to build capacities to enhance evidence use in very challenging political contexts. We hope that the lessons we have documented from BCURE's hard-won experience will enable future programmes to develop a deeper understanding of the broader contextual factors and multiple mechanisms that need to work together to create and reinforce evidence use, so that future capacity support has a stronger potential to catalyse self-transformation and improved effectiveness in government partners.

9. Recommendations for future programmes

This section builds on the conclusions, providing some practical pointers for future programmes. These are designed as a prompt for implementers to help take the insights from BCURE into account, rather than a comprehensive guide on how to design an effective capacity building programme. The six recommendations correspond to the six lessons in Section 8.

1. When choosing an entry point…

- **Identify a sector, institution or policy area where there is existing interest in evidence and clear incentives for reform,** such as opportunities around devolution, large-scale donor-supported programmes bringing a focus on evidence to underpin effectiveness, or national reforms leading to demands for results and M&E. Look out for senior individuals who may be 'champions' for the programme – people who are genuinely interested in reform and have the charisma and connections to open doors and bring others on board. However, do not rely on specific individuals too much, given the risk of staff turnover.
- **Identify a unit that can provide an entry point for capacity building, and consider its status within the ministry or wider government setting.** This might be a research or policy unit, which has the potential to become an internal advocate for evidence. It is important to consider the level of influence it has over policy and decision making, and whether it has any formal or informal authority to promote compliance with new evidence-informed approaches. If influence is low, for example where a unit is relatively new or has few staff, programmes should consider how their support could positively raise the profile and influence of the unit as an advocate for evidence within the setting. It is also important to think about the potential for the unit to mobilise additional resources that might allow reforms to be institutionalised beyond the lifespan of the programme, or enable targeted individuals to cascade their learning to other relevant staff, especially in government contexts where resources for implementation are scarce or donor-dependent.
- **Explicitly consider concerns around politicised decision making and corruption that set up powerful disincentives for EIPM.** Programmes may need to work to mitigate these, re-align incentives more positively to support evidence use, or identify pockets of good practice that can be built upon.

2. When seeking to accompany change, rather than impose it...

- **Invest in building trust, taking care to promote ownership and providing support in a collaborative way.** Effective accompaniment means avoiding falling into the role of service provider or consultant, ensuring that government stakeholders maintain technical involvement and do not over-delegate to the partner. This requires a clear focus on the 'learning-by-dong' objectives and a flexible but negotiated approach to specific tasks which emerge.
- **Monitor shifting incentives, both positive and negative, and assess how these might create new opportunities or challenges.** For example, a high-profile scandal might lead to the appointment of a new senior stakeholder who may not be supportive of EIPM, but equally may lead to increased scrutiny from donors with higher requirements for evidence.

- **Consider how the programme can be designed with sufficient flexibility to allow it to respond to emerging opportunities and challenges**, given the volatility of change in government settings. This requires the active engagement of both implementers and funders, as discussed in Section 4.3.

3. When attempting to promote change at an individual level, through catalysing self-efficacy, facilitation and reinforcement…

- **Align training or other activities targeting individuals with incentives in the broader environment.** A critical factor is the extent to which managers and senior staff encourage and support trainees to apply their learning, which in turn is related to broader incentives for or against evidence use within the civil service, when weighed up against partisan policymaking or widespread corruption. If senior managers are not proactively engaged, they may not allow trainees time to apply skills or try out new approaches, nor be receptive to new insights or challenges to decisions that trainees may bring as a result of a more evidence-informed approach. Incorporating soft skills as well as technical skills in training is one way to help trainees engage with the broader environment – around how to communicate and present evidence, use evidence to influence decisions, understand the needs of evidence users or negotiate for resources. Working in parallel through a 'top down, bottom up' approach may help ensure that wider 'top down' incentives to adopt new ways of working are in place, so that people reached through 'bottom up' training approaches actually use their new skills.
- **Follow the principles of adult learning theory when designing activities.** See Spotlight 2 for more details. This includes ensuring training is closely targeted to those who can apply it, is directly relevant to their day-to-day work, and that trainees have immediate opportunities to apply their skills. This also suggests going beyond one-off training to provide follow-up and on-the-job support to help trainees use and embed their skills over time. Ensure training is practical and participatory, using local case studies or live policy examples, and that facilitators are knowledgeable, patient and confident, and understand the specific sector as well as the broader national context. Making training participatory is not always easy in contexts where training is typically fairly didactic, and may require significant capacity support to local trainers.

4. When seeking to catalyse a 'critical mass'…

- **Consider how to create incentives and harness resources for cascading.** For trainees to share their new EIPM skills there need to be organisational incentives to use evidence, and senior management needs to be sufficiently engaged to give time, resources and support to trainees to share their learning with colleagues.
- **Building in an explicit 'training of trainers' strategy, supported by a 'clustering' approach, may help trainees develop the social connections to provide mutual support.** There is limited evidence on this from BCURE, but some suggestion that clustering a group of trainees within the same unit may allow them to act as a 'focal point' for promoting new behaviours.

5. When aiming to showcase the value of evidence through policy pilots or evidence tools…

- **Identify a recognised policy problem, so that an evidence-informed approach can be showcased while meeting priority policy objectives.** This may be a contested policy problem which could benefit from an evidence-based analysis involving key stakeholders, or a recognised problem in service delivery that a data tool could help address.
- **Seek to capitalise on existing work and partnerships, identify allies, and leverage external resources.** For example, previous research and collaborations might provide an entry point. Opportunities may not present themselves immediately, but rather emerge over time as relationships and trust are established through earlier activities.
- **Ensure the right people are involved from within and outside government.** High-level support is essential to the success of the process, ensuring that activities are given priority, as they require significant time and cooperation. Technical working groups require the right mix of technical and academic stakeholders, alongside senior managers with sufficient authority to underscore commitment to the process, but not so senior to engage in hands-on work.
- **Provide flexible, responsive, hands-on support to 'co-produce' policies and tools**, rather than having external consultants develop them. Co-production requires staff with strong technical and interpersonal skills, including responsiveness and commitment, 'going above and beyond' to support the process, follow-up on tasks, coordinate, and 'push things along' to ensure things get done; all underpinned by a practical understanding of policymaking processes and local realities, alongside the ability to convene stakeholders and the credibility to negotiate priorities between different groups.
- **Synergise with other programme activities.** A 'co-production' approach should provide opportunities for the officials involved to develop or deepen their skills in accessing, appraising and applying evidence, helping to embed the learning from EIPM training courses through 'on-the-job' learning. There may also be opportunities to produce or pilot guidelines or other evidence tools as part of the process.
- **Remain alert as to how a new tool may affect incentives, positive or negative, to use evidence.** Just because good quality tools are there does not mean they will be used, especially when there are powerful incentives to make decisions based on factors other than data. It is important to be aware of where evidence and data may challenge the status quo, disrupt practices that could lead to personal gain, or put an individual's role in jeopardy by highlighting inefficiencies.

6. When attempting to promote the longer-term adoption of reforms into government systems…

- **Be clear from the design phase what the ultimate goal is.** The aim should not be to simply successfully trial a new process, tool or practice, but to promote its formal adoption at an organisational level, and ensure it is supported, resourced and incentivised by senior managers – in order for this in turn to continue catalysing, deepening and incentivising individual-level change.

- **Link activities to ongoing government initiatives and resources which may provide opportunities for adoption and scaling.** To move beyond interesting and 'nice to have' examples of how evidence may be used to greater effect in decision making, activities should be linked into wider initiatives for reform and, ideally, engage partners with further networks and resources available to them to facilitate the adoption and scale-up of pilot initiatives.
- **Explicitly identify the political economy risks that might skew genuine change into isomorphic mimicry**, and aim to continuously manage this risk every step of the way. 'Genuine' adoption requires ongoing government ownership and resources to be catalysed beyond the project, as high-level incentives shift and new opportunities rise and fall in dynamic political environments.

10. References

Andrews, M. (2013). The limits of institutional reform in development: Changing rules for realistic solutions. Cambridge University Press. Link.

Andrews, M., Pritchett, L. and Woolcock, M. (2015). Doing Problem Driven Work. Harvard Kennedy School Faculty Research Paper. Link.

Andrews, M., Prichett, L. and Woolcock, M. (2016). Doing Iterative and Adaptive Work. CDI Working Paper. Link.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2): 191–215. Link.

Bandura, A. (1988). Organizational applications of social cognitive theory. *Australian Journal of Management, 13*(2): 275–302. Link.

Booth, D. and Unsworth, S. (2014). Politically Smart, Locally Led Development. ODI Discussion Paper. Link.

Bryan, R.L., Kreuter, M.W. and Brownson, R.C. (2009). Integrating adult learning principles into training for public health practice. *Health promotion practice, 10*(4): 557–63. Link.

Cairney, P. (2016). *The Politics of Evidence-Based Policy Making*. London: Palgrave Macmillan UK. Link.

Cummings, C. (2015). Fostering innovation and entrepreneurialism in public sector reform. *Public Administration and Development, 35*(4): 315–28. Link.

DFID. (2010). How To Note on Capacity Building in Research. DFID Practice Paper. Link.

Dunn, L. (2002). Theories of Learning. Learning and Teaching Briefing Paper Series, Oxford Centre for Staff and Learning Development, (Oxford Brookes University). Link.

du Toit, A. (2012). Making Sense of 'Evidence': Notes on the Discursive Politics of Research and Pro-Poor Policy Making. Working Paper 21, Institute for Poverty, Land and Agrarian Studies (PLAAS). Link.

Eden, D. and Avirma, A. (1993). Self-efficacy training to speed reemployment: Helping people help themselves. *Journal of Applied Psychology, 78*: 352–60. Link.

FAO. (2010). Capacity Development Learning Module 1. Link.

Faustino, J. and Booth, D. (2014). Development Entrepreneurship: How Donors and Leaders can foster Institutional Change. Working Politically in Practice Series. Link.

Fishbein, M. and Middlestadt, S. (1994). Noncognitive effects on attitude formation and change: Fact or artifact? *Journal of Consumer Psychology, 4*: 181–202.

Funnell, S.C. and Rogers, P.J. (2011). Purposeful Program Theory: Effective Use of Theories of Change and Logic Models. San Francisco, CA: Jossey-Bass.

Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P. and Kyriakidou, O. (2004). Diffusion of innovations in service organizations: systematic review and recommendations. *The Milbank Quarterly, 82*(4), 581–629. Link.

Krause, P. (2013). Of institutions and butterflies: is isomorphism in developing countries necessarily a bad thing? ODI Background Note. Link.

Leavy, J., Boydell, E. and Mcdowell, S. (2017). Making Progress: BRACED at the mid-term. Link.

Lemire, S.T., Nielsen, S.B. and Dybdal, L. (2012). Making contribution analysis work: A practical framework for handling influencing factors and alternative explanations. *Evaluation*, 18(3). Link.

Liverani, M., Hawkins, B. and Parkhurst, J.O. (2013). Political and institutional influences on the use of evidence in public health policy. A systematic review. *PloS one, 8*(10). Link.

Lockwood, C. and Pearson, A. (2013). A comparison of meta-aggregation and meta-ethnography as qualitative review methods. Wolters Kluwer/Lippincott Williams & Wilkins. Link.

Lyon, A.R., Stirman S.W., Kerns, S.E. and Bruns, E.J. (2011). Developing the mental health workforce: Review and application of training approaches from multiple disciplines. *Administration and Policy in Mental Health, 38*(4): 238–53. Link.

Mayne, J. (2011). Addressing cause and effect in simple and complex settings through contribution analysis. In R. Schwartz, K. Forss and M. Marra (eds), *Evaluating the Complex*. Piscataway, NJ: Transaction. Link (draft).

Mayne, J. (2012). Contribution analysis: Coming of age? *Evaluation, 18*(3), 270–80. Link.

Nutley, S., Walter, I. and Davies, H.H.T. (2007). *Using Evidence: How Research Can Inform Public Services*. Bristol: The Policy Press. Link.

Oliver, K., Lorenc, T. and Innvær, S. (2014). New directions in evidence-based policy research: A critical analysis of the literature. *Health research policy and systems/BioMed Central, 12*, p. 34. Link.

O'Neil, T. (2016). Using adaptive development to support feminist action. ODI Report, (March). Link.

Pawson, R. and Tilley, N. (1997). *Realistic Evaluation*. London: Sage. Link.

Pawson, R. and Tilley, N. (2004). Realist Evaluation. Link.

Punton, M., Brown, C., Echt, L., Fisher, C., Lloyd, R., Morton, S. and Vogel, I. (2016a). How Can Capacity Development Promote Evidence-Informed Policymaking? Literature Review for the BCURE programme. Brighton. Link.

Punton, M., Vogel, I. and Lloyd, R. (2016b). Reflections from a realist evaluation in progress: scaling ladders and stitching theory. *CDI Practice Paper*, (18). Link.

Ramalingam, B. (2013). Aid on the Edge of Chaos: Rethinking International Cooperation in a Complex World. OUP Oxford. Link.

Rogers, E.M. (2003). *Diffusion of Innovations*. 5th edition New York: The Free Press. Link.

Shaxson, L. (2014). Investing in Evidence: Lessons from the UK Department for Environment, Food and Rural Affairs (Working Paper No. 2). Australian Aid Knowledge Sector Initiative. Link.

Shaxson, L. (2016). Lessons for Building and Managing an Evidence Base for Policy. The Knowledge Sector Initiative. Link.

Smith, K.E. and Joyce, K.E. (2012). Capturing complex realities: Understanding efforts to achieve evidence-based policy and practice in public health. *Evidence & Policy: A Journal of Research, Debate and Practice, 8*(1), 57–78. Link.

Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R. and Befani, B. (2012). Broadening the range of designs and methods for impact evaluations. *DFID Working Paper, 38* (April 2012). Link.

Valters, C. (2015). Theories of change: time for a radical approach to learning in development. London. Link.

Waldman, T. (2014). The Use of Statebuilding Research in Fragile Contexts: Evidence from British Policymaking in Afghanistan, Nepal and Sierra Leone. *Journal of Intervention and Statebuilding, 8*(2–3): 149–72. Link.

Walter, I., Nutley, S. and Davies, H. (2005). What works to promote evidence-based practice? A cross-sector review. *Evidence & Policy: A Journal of Research, Debate and Practice, 1*(3): 335–64. Link.

Westhorp, G. (2014). Realist Impact Evaluation: An Introduction (ODI Methods Lab). London. Link.

Wimbush, E., Montague, S. and Mulherin, T. (2012). Applications of contribution analysis to outcome planning and impact evaluation. *Evaluation, 18*(3), 310–29. Link.

Wong, G., Westhorp, G., Manzano, A., Greenhalgh, J., Jagosh, J. and Greenhalgh, T. (2016). RAMESES II reporting standards for realist evaluations. *BMC Medicine, 14*(1), 96. Link.

Wong, G., Westhorp, G., Pawson, R. and Greenhalgh, T. (2013). Realist Synthesis: Rameses Training Materials. London: Health Research Health Services and Delivery Research Program. Link.

World Bank. (2017). World Development Report 2017: Governance and the Law. Washington, DC: World Bank. Link.

Yanguas, P. (2017). Varieties of state-building in Africa: Elites, ideas and the politics of public sector reform. ESID Working Paper. Link.

Yin, R.K. (2003). *Case Study Research: Design and Methods*. Thousand Oaks, CA: Sage. Link.

We want the resources invested in international development to have the greatest possible impact on people's lives. We provide the insight and ideas to ensure that they do.

Itad
Preece House
Davigdor Road
Hove BN3 1RE
United Kingdom

+44 (0) 1273 765 250

itad.com

Annexes for the Final Evaluation of the Building Capacity to Use Research Evidence (BCURE) programme

Contents

1 Terms of reference and BCURE logframe

The original Terms of Reference and BCURE Logframe are presented below. Note that the project end-date moved to November 2017 as a result of various extensions, in order to allow sufficient time for projects to complete their activities.

ITT Volume 3

Terms of Reference for Evaluation of Approaches to Build Capacity for Use of Research Evidence

Title:	Evaluation of Approaches to Build Capacity for Use of Research Evidence

A. Introduction

1. DFID is committed to supporting research and its effective use by policymakers and practitioners. This commitment is driven by the assumption that making more effective use of evidence will enable countries to make better policy and programme decisions, ultimately enabling them to develop more rapidly and sustainably. In the past DFID has focused on the supply of high quality research, with less work done to ensure that there is a corresponding demand for research evidence in developing countries. However, emerging evidence suggests that there are significant gaps in capacity of decision makers in the South to use research effectively, which is hampering research uptake.
2. In response to these gaps, DFID has recently launched a programme called Building Capacity to Use Research Evidence (BCURE). This is a three-year £13 million programme aimed at increasing the ability of policymakers, practitioners and research intermediaries in the South to use research evidence for decision making. The overall goal of the BCURE programme is for 'Poverty reduction and improved quality of life', and its overall purpose is for 'Policy and practice to be informed by research evidence'.
3. Improving the use of research evidence in decision making is a relatively new area for donor support, meaning that the evidence base on what works is limited. Therefore, a significant component of the BCURE programme is an evaluation of both – the wider challenge of supporting evidence-based decision making and the value of the BCURE programme itself, drawing comparisons to other capacity-building programmes where appropriate. In doing so, the primary objective of the evaluation is to **help strengthen the global evidence base on whether capacity-building approaches to supporting evidence-informed policymaking can be a cost effective way to reduce poverty and, if so, how can they be implemented to achieve the greatest impacts.**
4. The direct recipients of the services will be DFID's Research and Evidence Division and governance cadre. The published final report is expected to be of value to donors and practitioners in the research uptake community.

B. Building Capacity for the Use of Research Evidence (BCURE)

5. The BCURE programme was procured in 2012/2013 through open competition. A large number of initial proposals were received, of which 12 were selected to develop into full proposals, including theories of change, work plans and logical frameworks. Of these 12 proposals, five were selected for funding and have now progressed to the contracting stage. A sixth proposal is still under discussion.
6. Each of the five successful proposals will employ a different approach to capacity building. The five projects will begin between September 2013 and January 2014, last three years each and end between August and December 2016. Each project is worth between £1.3 and £3.4 million. Three of the projects have already been issued contracts, with the remaining two projects expected to receive contracts within the next month.

	Primary provider	Description	Focus countries
A	Adam Smith International	Support African cabinets to implement evidence-based decision processes, focusing on post-conflict states	Sierra Leone, Liberia and South Sudan
B	Finalising contract	African-led programme to strengthen use of research evidence for health policymaking	Kenya and Malawi
C	Finalising contract	Develop online training on use of evidence aimed at policymakers	India, Pakistan and Afghanistan
D	INASP	Develop and implement courses on use of evidence, focusing on civil servants and parliamentarians	Ghana, Zimbabwe and South Africa
E	University of Johannesburg	Develop and implement courses on evidence, focusing on civil servants	South Africa and Malawi

7. A decision will be made on whether to progress with the sixth proposal shortly; further details on this proposal may therefore be shared with those bidders invited to progress to the ITT stage.
8. A short overview of each project is provided in Annex 1. The full project proposals will be will be shared with those invited to submit a full tender. The BCURE programme business case and intervention summary provides further background to the overall programme design, including the original theory of change. It can be accessed on the project pages of DFID website. This ToR should be considered as DFID's definitive thinking on this evaluation, rather than the BCURE business case.

C. Purpose, scope and evaluation questions

9. The primary purpose of this evaluation is to 'strengthen the evidence base to support evidence-informed policymaking in developing countries'. This assessment will help DFID and others make better choices in the future, when deciding whether and how to support and implement capacity-building programmes on evidence use. In order to make this assessment, the evaluation is expected to draw on both the BCURE programmes and the existing body of evidence related to building capacity to use evidence for decision making.
10. The secondary purpose of this evaluation is to 'evaluate the success and value for money of the BCURE projects in building capacity to use research evidence for decision making'. This assessment will help inform DFID decisions about whether to provide additional funding to these projects beyond the original three-year contract.
11. The provisional evaluation questions are:

i) What different factors influence the extent to which policymaking organisations in developing countries use research evidence for decision making?

- What organisational structures, processes and systems help or inhibit the use of evidence by policymaking institutions?
- What characteristics help or inhibit the use of evidence by individuals within those organisations? Including (but not limited to):
  - Educational history (including subject focus, level of attainment, location of education, predominant pedagogical approach, etc.)
  - Existing skills or knowledge
  - Cultural or attitudinal behaviour
- What wider institutional factors support or inhibit the use of evidence by policymaking institutions, including the role of civil society?

ii) How effective are the BCURE projects in achieving their stated outcome of increasing the use of research evidence in decision making?

- In each project, what were the observable changes in …
  - organisational policies, systems or process;
  - individuals' knowledge and skills;
  - the wider institutional environment (including civil society);

… and how effective were these in increasing the use of research evidence in decision-making processes?

- To what extent were these changes driven through local leadership/ownership (i.e. how endogenous was the process) and what effect did this have on the projects' effectiveness?
- What is the relative quality of support provided by the project when designing and implementing changes to organisational policies, systems and processes? Including (but not limited to):
  - How well did this support and the final changes meet organisational needs? (i.e. to what extent did the projects implement a 'best fit' approach?)
  - What is the likely medium and long-term sustainability of these changes?
- What is the relative quality of training and pedagogy in the capacity-building approach adopted by each project? Including (but not limited to):
  - To what extent to the pedagogical approaches used match with 'best practice' for supporting adult and organisational learning?
  - How well does this support meet individual learning needs? (i.e. to what extent did the projects implement a 'best fit' approach?)
- What approaches are most effective in building the capacity of local civil society organisations? Including (but not limited to):
  - How effectively did the projects increase the capacity of local civil society organisations to use effective pedagogical approaches in training?
  - How effective were multi-country networks in increasing the local capacity of civil society organisations?
- Overall, how does each project's model of capacity building relate to other models of capacity building – both within and outside of the BCURE programme – in terms of value for money?

iii) Drawing on the lessons from the BCURE programmes and other relevant interventions, what factors influence the effectiveness of capacity-building interventions in increasing the use of research evidence?

- What organisational-level changes introduced by capacity-building interventions are most effective at increasing the use of research evidence in a policymaking institution?
- What programmatic factors help or inhibit the uptake of these changes? Including (but not limited to):
  - Which roles in an organisation should capacity-building interventions target, in order to maximise the uptake of evidence in decision making?
  - How should senior decision makers be involved in designing and/or overseeing capacity-building interventions?
  - How can organisational-level changes best help support efforts to increase individual capacity to use research evidence and vice versa?
- What programmatic factors influence how effective capacity-building interventions are at increasing an individual's ability to use research evidence effectively? Including (but not limited to)
  - What pedagogical approaches to increasing individual capacity to access, appraise and use research evidence are most effective in increasing objectively measured capacity?
  - Looking at different types of capacity building (e.g. training, mentoring, secondments etc.) what features predict success in increasing individual capacity to use research?
- To what extent can a capacity-building programme influence the wider institutional environment, in order to help support the greater uptake of research evidence in decision making? Including (but not limited to)
  - How effective are efforts to strengthen civil society networks in supporting greater uptake of research evidence?

- What factors are important for the long-term sustainability of changes implemented by capacity-building interventions? Including (but not limited to)
  - To what extent do changes in individual capacity affect the overall culture of evidence use in a policymaking institution?

iv) What impacts do capacity-building interventions that are specifically aimed at increasing the use of research evidence have on …

- Increasing the use of research evidence in actual policy and programme decision making?
- Improving the relative quality of policies and programmes, in comparison with other technical assistance programmes aimed at improving policymaking and/or supply-side research evidence interventions?[1]

12. In order to answer these questions, it is expected that the evaluation will develop a methodology or framework for measuring the degree to which research evidence has been used in policymaking process.
13. There is some scope to amend or add to evaluation questions. Short-listed bidders will be invited to suggest what (if any) changes that they would make to the evaluation questions, as part of the ITT. Further guidance on this may be provided in the ITT pack.

D. Design and methodology

14. Those tenderers invited to submit a full tender are invited to propose an evaluation design and methodology that best delivers the purpose and required outputs. This should also cover the potential risks and challenges for the evaluation and how these will be managed. DFID has not endorsed particular methodology(ies) for the conduct of research on capacity-building programmes. We would expect a design that takes a mixed methods approach, combining primary data collection from the BCURE projects and secondary evidence synthesis and analysis from existing sources. Primary data collection in non-BCURE countries and/or interventions may be proposed.
15. Tenderers should spell out with the approach and methods which they will use. It would be helpful if bidders explain why they selected the options they propose to use and briefly outline what other options they considered, if any. Please note that we are committed to quality and rigour in line with international good practice in evaluation.
16. The successful tenderer will refine their proposal within the first six months of the contract, in consultation with DFID, the BCURE project providers and other relevant stakeholders.
17. Proposed designs should clearly show how they will address well-known challenges with evaluating the impact of capacity-building programmes aimed at long-term cultural and institutional changes. These challenges will include:
    - Complexity and time lag: The pathway from increased beneficiary skills/knowledge to embedded changes in practice can be long and complex. In addition, the duration between 1) beneficiaries acquiring new skills and/or knowledge, 2) the application of these skills when designing policies and programmes, and 3) benefits to poor people from improved policies can be long and variable, and may be outside the span of this evaluation. While these two challenges affect all evaluations of capacity-building programmes, they are particularly relevant to this evaluation because the BCURE projects are being implemented simultaneously with (rather than preceding) the evaluation. This means that the proposed designs should acknowledge the degree to which they expect to be able to answer the evaluation questions within the timeframe.
    - Contribution/attribution: the BCURE capacity-building support may well not be the only factor impacting on the changes observed.
    - Context: the evaluation will need to draw lessons from across a wide range of countries and contexts.

[1] Technical Assistance programmes could include sector or organisation specific support aimed at improving the relative quality and/or effectiveness of programmes or policies. Supply side research evidence interventions refer to support to online research portals and other research uptake activities.

18. The evaluation is expected to focus on the use of research evidence in a broad sense, i.e. published academic research papers; statistical databases; 'established' (i.e. widely debated and accepted) policy papers and positions; and evaluation findings. It does not include experiential evidence (i.e. evidence based on professional insight, skills or experience) or all types of contextual evidence (i.e. evidence based on likely uptake or impact within a given community), though some type of contextual evidence may be usefully included. Tenderers are welcome to include a definition of research evidence in their proposals, where they feel this may be helpful to clarify their proposed research design and approach.

Specific requirements: evaluation design

19. The evaluation must include the development of a programme-level theory of change (ToC) during the inception phase. While we have not taken a view on the whether this ToC should or should not have a central role in the evaluation approach and analysis, this will be a valuable tool for DFID and other organisations considering designing or funding similar types of capacity-building programmes. At a minimum, this ToC should draw upon the initial theories of change presented in the BCURE business case and the five BCURE project proposals.
20. The evaluation should include at least one case study per BCURE project.
21. Secondary evidence synthesis and analysis should be conducted in line with DFID's guidance on 'Assessing the Strength of Evidence' (2013). The literature review should include an examination of the different analytical frameworks used to evaluate capacity for use of research evidence.

Sources

22. Sources of data that will be used in the evaluation would, at a minimum, include:
    - **Background documentation:** BCURE business case and project proposals.
    - **Secondary data and literature:** a document review and analysis of existing evidence. This should include research evidence on interventions to build capacity to use evidence. Research/evaluations carried out in low income contexts will be particularly relevant, though tenderers should also consider what lessons can be drawn from research carried out in other contexts. The analysis may also draw relevant lessons from research on related themes – for example research into effective approaches to supporting adult learning or research into organisational learning and change.
    - **Primary data gathered by the evaluation team:** e.g. interviews with key partners and users – including face-to-face meetings – surveys or other data collection methods with beneficiaries and stakeholders.
    - **Primary data gathered by the BCURE project providers:** e.g. data from the projects' monitoring frameworks, progress reporting etc.

In choosing an approach and methods, the tenderer should as far as possible, set out the different data sources they expect to use – including types of primary data – and what weighting they would expect to attribute to data when forming their evaluation conclusions.

23. The BCURE projects will be an important source of data. The evaluation is therefore expected to work closely with BCURE project providers, in order to:
    - Support providers to suggest amendments to their draft monitoring frameworks, in order to maximise alignment with the evaluation objectives;
    - Comment on monitoring tools developed by providers, such as training assessment forms, and the information gathered from those tools; and
    - Participate in annual BCURE lesson learning meetings.
24. BCURE projects were made aware in advance of DFID's plans for independent external evaluation; good levels of cooperation can be anticipated with regard to reasonable requests to support the evaluation. Input from projects does not need to be costed.

25. Noting the volume and quality of applications to the BCURE programme, tenderers invited to submit an ITT may wish to suggest a role within the evaluation for certain unsuccessful applicants (of full proposals and/or concept notes). Further information on this will be included in the ITT information pack.

Ethics

26. The evaluation should ensure that it adheres to the ethical evaluation policies of DFID and the evaluation principles of accuracy and credibility.

E. Timing and Scope

27. The evaluation should start as soon as possible, in order to facilitate early engagement with BCURE projects. Taking into consideration logistical and procurement requirements, our anticipated start date is around April 2014. The evaluation will last approximately three years and three months (39 months), ending mid-2017. However, bidders may suggest a later completion date in 2017, where they believe that this will significantly strengthen the evaluation findings, given their research design. There is the option of a one-year extension in case of unforeseen circumstances, though DFID's strong preference is for the evaluation to conclude no later than December 2017.
28. DFID also reserves the right to scale up/scale back the evaluation programme depending on the requirements.

29. The evaluation is expected to include some assessment of project activities in all 11 of the BCURE beneficiary countries. We do not have a view as to what level of engagement in each country would be most appropriate, nor whether engagement should be split equally between all countries or focus on particular countries. The successful provider will be responsible for arranging their own logistical arrangements. However, the BCURE project providers will provide some support with identifying and contacting key contacts.
30. The primary focus of this evaluation is approaches to increase the systematic use of research evidence to inform policymaking. Efforts to *influence* particular policies with a given piece of research are not the focus of this evaluation. Tenderers are welcome to include a definition of 'policies' in their proposals, where they feel this may be helpful to clarify their proposed research design and approach.
31. Capacity building/development refers to the capacity of individuals, organisations and the broader institutional framework within which individuals and organisations operate to deliver specific tasks and mandates.
32. The evaluation is expected to focus on Lower-Income Countries and those middle-income countries with a high poverty burden. However, the evaluation may consider evidence from other countries where this is helpful.

F. Outputs

33. The evaluation team will produce the following outputs:

- **Inception Report and initial literature assessment** within six months. This should include refinements/amendments of evaluation questions and full methodology; overarching theory of change; suggested amendments to the monitoring frameworks for the BCURE projects; identified sources of data and risk management strategy; communications strategy; work plan and any proposed budget revisions (within the agreed total contract value).
- **Stage 1 of the evaluation** within 12 months**,** comprising findings from secondary data and initial collection of primary data. This report should focus on evaluation question 1, though may helpfully include findings for the other evaluation questions, as available.
- **Stage 2 of the evaluation** by April 2016, comprising an initial report on evaluation question 2, in order to inform decisions on future DFID support under the BCURE programme. The exact format for Stage 2 will be agreed during the inception phase. As the projects will have only completed between 28 and 32

months of their 36 month contracts, this will impose some constraint on the strength of conclusions possible at this stage.

- **Draft Stage 3 of the evaluation** within 36 months (approximately December 2016), comprising a draft report of all the evaluation questions. This report will be commented on by DFID, with areas for revision and further research highlighted.
- **Final Stage 3 of the evaluation** within 39 months, comprising the full report (maximum of 150 pages with a maximum six-page Executive Summary) that incorporates feedback obtained on the draft report. This report will be externally peer reviewed, to be organised by DFID.
- Appendices with details on the methodology, informants, etc.

34. DFID's intention is for the evaluation findings to be available and shared widely within the international community, in order to strengthen the evidence base in this area. This means that publication of the evaluation findings – in particular, Stages 1 and 3 – will be required to comply with DFID's Enhanced and Open Access Policy. In addition, tenderers are invited to suggest how they would share findings through peer reviewed publications and other communication outputs and channels, as part of the ITT.

G. Management, reporting and financial arrangements

Management arrangements

35. The evaluation will be overseen by a steering group, who will be responsible for approving the evaluation outputs and commenting on draft reports. The steering group shall comprise:
    - Jessica Prout and Nathanael Bevan from DFID's Evidence into Action team, who are managing the BCURE programme
    - A DFID evaluation adviser and/or governance specialist not directly involved in BCURE
    - One or two external representatives
36. Day-to-day management of the study will be undertaken by Jessica Prout and the deputy programme manager of the Evidence into Action team.

Financial and reporting arrangements

37. Bidders are invited to explain how they would link payment to results, as part of the ITT. DFID's preference would be for payment to be made against achievement of quarterly or bi-annual milestones, as a form of output-based contract. Payments must be accompanied by short technical reports, detailing progress against the milestones, work plan and budget.
38. In addition to technical reports, the successful bidder is expected to meet bi-annually with the steering group. As part of these meetings, they will be expected to deliver up to four presentations to the steering group (one in presenting the inception report; one in presenting Stage 1; one in presenting Stage 2; and one in presenting the draft Stage 3 report). Meetings at which the successful bidder is presenting will take place in London; other meetings will take place either in London or via telephone, depending on logistics.
39. Mandatory financial reports include an annual forecast of expenditures (the budget) disaggregated monthly for the financial year April to March. This should be updated either quarterly or bi-annually, in line with the agreed payment schedule, alongside a report of actual expenditure over the period. The successful bidder must also submit yearly external audit reports on their annual financial statements.
40. Key performance indicators (KPIs) will be agreed with the successful bidder during the inception phase.

Inception phase

41. The evaluation will have an inception phase of up to eight months, during which the inception report and initial literature will be finalised, submitted to and agreed by DFID. There will be a formal contract break at the end of the inception phase and DFID reserves the right to terminate the contract at that point if the work undertaken during the inception phase is unsatisfactory or agreement cannot be reached on the remainder of the evaluation (budget / detailed methodology and work plan).

H. The evaluation team

42. Pre-Qualification Questionnaires (PQQ) from suitably qualified organisations and consortia are equally welcome. Lead organisations for the consortia contracted to deliver the BCURE projects are not eligible to apply (as set out in 41. in the BCURE Terms of Reference). Other BCURE consortium members are eligible to apply, but must fully explain in an Annex to their PQQ how they would manage any conflict of interest that may potentially arise. The proposed evaluation team may not include any individual who is contracted as part of a BCURE project.
43. The supplier will design, co-ordinate and draw together the evaluation findings in a final report. They will quality assure the outputs and validate the data collected.
44. The BCURE project providers will also seek to facilitate access to stakeholders who have direct links with the programme, but the evaluation team will have to make direct approaches to other stakeholders and beneficiaries who are in scope of their evaluation design.
45. DFID welcomes proposals that:
    - Where the evaluation is being conducted by one organisation from a high income country, includes plans in the PQQ for helping to build local capacity to conduct high quality evaluations.
    - Where the evaluation is being conducted by a consortia, that this either includes member organisations from low or middle-income countries (preference), or includes plans in the PQQ for helping build local capacity to conduct high quality evaluations.

Skills and qualifications

46. As outlined in the PQQ, the essential competencies and experience that the contractor will need to deliver the work are:
    - Extensive knowledge and application of evaluation methods and techniques, preferably with experience in implementing evaluations of a similar scope and size to this ToR
    - Strong qualitative and quantitative research skills
    - A good understanding of capacity building
    - Strong analysis, report writing and communication skills, preferably with experience in publishing evaluation and/or research findings in peer reviewed publications
    - Experience of engaging with Southern partners
47. Desirable competencies and experience are:
    - Experience in evaluating, research or delivering capacity-building interventions
    - A good understanding of research uptake
    - Expertise in assessing value for money

Further advice

48. Enquiries regarding these Terms of Reference can be submitted as dialogue questions via the DFID supplier portal. Where appropriate, answers to these questions will be posted and will be visible to all potential suppliers.

Duty of Care

49. The supplier will be responsible for the safety and well-being of their personnel and Third Parties affected by their activities, including appropriate security arrangements. They will also be responsible for the provision of suitable security arrangements for their domestic and business property. The supplier is responsible for ensuring that appropriate arrangements, processes and procedures are in place for their personnel, taking into account the environment they will be working in and the level of risk involved in delivery of the Contract (such as working in dangerous, fragile and hostile environments, etc.). The supplier must ensure their personnel receive the required level of training and where appropriate complete a UK government approved hostile environment or safety in the field training prior to deployment.

50. Tenderers must develop their PQQ Response and Tender (if Invited to Tender) on the basis of being fully responsible for Duty of Care. They must confirm in their PQQ Response that:
    - They fully accept responsibility for Security and Duty of Care.
    - They understand the potential risks and have the knowledge and experience to develop an effective risk plan.
    - They have the capability to manage their Duty of Care responsibilities throughout the life of the contract.

    If you are unwilling or unable to accept responsibility for Security and Duty of Care as detailed above, your PQQ will be viewed as non-compliant and excluded from further evaluation.

51. Acceptance of responsibility must be supported with evidence of Duty of Care capability and DFID reserves the right to clarify any aspect of this evidence. In providing evidence, interested suppliers should respond in line with the Duty of Care section in Form E of the PQQ.
52. DFID will provide risk assessments for the relevant countries when issuing the ITT pack. Bidders will be expected to prepare Duty of Care plans as part of their technical response.

I. Budget

The budgeted expenditure for this work over a three-year period is between £700,000 and £950,000.[2] Value for money will be a key criterion in selection and the final budget will be agreed with the successful provider.

[2] The BCURE business case budgeted for up to £2 million to be split between three evaluations on research capacity building and uptake.

BCURE Joint Logframe

PROJECT NAME	Building Capacity to Use Research Evidence (BCURE) programme						
IMPACT	**Impact Indicator 1**		**Baseline**	**Milestone 1**	**Milestone 2**	**Target (date)**	
Better design and implementation of government programmes and policies leads to reduced poverty	Worldwide governance indicator on government effectiveness	**Planned**	From 2012 dataset, listing by rank: South Sudan: 3 Afghanistan: 7 Zimbabwe: 11 Sierra Leone: 11 Liberia: 12 Bangladesh: 22 Pakistan: 23 Kenya: 35 Malawi: 38 India: 47 Ghana: 52 South Africa: 64				
		Achieved					
			Source				
	Impact Indicator 2		**Baseline**	**Milestone 1**	**Milestone 2**	**Target (date)**	
	Inequality-adjusted Human Development Index (IHDI)	**Planned**	From 2012 dataset, listing by IDHI score South Africa: 0.629 Ghana: 0.558 India: 0.554 Kenya: 0.519 Bangladesh: 0.515 Pakistan: 0.515 Malawi: 0.418 Zimbabwe: 0.397 Liberia: 0.388 Afghanistan: 0.374 Sierra Leone: 0.359				

			South Sudan: unranked				
		Achieved					
			Source				

OUTCOME	**Outcome Indicator 1**		**Baseline**	**Milestone 1**	**Milestone 2**	**Target (date)**	**Assumptions**
Strengthened and embedded in-country capacity (skills, systems and culture) to access, appraise and apply research evidence and data, which influences international best practice.	Changed skills and/or processes in partners have led to an increased use of evidence in policy and programme decision making, as detailed in case studies (cumulative)	**Planned**	No data available	Six case studies (one per project)	12 case studies (two per project)	18 case studies (three per project)	Evidence-informed policy leads to better decision making and greater poverty reduction.
		Achieved					
			Source				
			Project reports, verified by DFID technical leads				
INPUTS (£)	**DFID (£)**		**Govt (£)**	**Other (£)**	**Total (£)**	**DFID SHARE (%)**	
						100%	
INPUTS (HR)	**DFID (FTEs)**						
	1.5						

OUTPUT 1	Output Indicator 1.1		Baseline	Milestone 1	Milestone 2	Target (date)	Assumption
Greater use of evidence in cabinet decision making in Africa, with a focus on Sierra Leone, Liberia and South Sudan (see nested logframe 1)	Cabinet secretaries have improved ability to oversee revised Cabinet processes, as measured by: - Revised Cabinet manuals are developed and used - Tracking systems developed and used to oversee implementation of Cabinet decisions - Number of trained policy analysts (or equivalent) in Cabinet Secretariats that are able to review evidence use - Proportion of strategic* proposals that are reviewed for quality by the Secretariats	**Planned**	* Cabinet manuals out of date * No effective process for tracking implementation * No policy analysts * No proposals reviewed by Cabinet Secretariat	* Revised cabinet manuals in Sierra Leone and Liberia * New tracking systems developed for monitoring cabinet proposals * At least 3 trained policy analysts in place over 3 countries * 15% of strategic proposals are reviewed	* Revised cabinet manual in South Sudan and support in place in Sierra Leone and Liberia * New tracking system approved and being used in all 3 countries * At least 6 trained policy analysts over 3 countries * 50% of strategic proposals are reviewed	* Cabinet Secretariat processes conducted in line with revised manuals * Cabinets have accurate data on implementation progress * At least nine trained policy analysts over 3 countries * 75% of strategic proposals are reviewed	1) Cabinet Secretariats have sufficient budgets and political backing to implement project activities 2) High-level support from Presidents and Ministers to agree and implement reforms, including providing the necessary staff time and resources from ministries 3) Cabinet Secretaries and other senior officials are available to participate in international workshops on given dates. 4) That political or other external events does not prevent programme implementation; in particular, that South Sudan remains stable enough to engage meaningfully in project
		Achieved					
		Source					
		Quarterly reports; Cabinet Secretariat monitoring tools and data; training records; discussions with beneficiaries					
	Output Indicator 1.2		**Baseline**	**Milestone 1**	**Milestone 2**	**Target (date)**	
	Ministers have greater ability to interrogate the quality of proposals submitted to Cabinet, as measured by: - Proportion of strategic* Cabinet proposals that are circulated to Ministers prior to Cabinet - Cabinet committee structures implemented - Proportion of relevant Cabinet items considered by Cabinet committee - Percentage of all Ministers who participate in workshops and describe it as 'good' or 'excellent' (cumulative)	**Planned**	* Between 0 and 15% compliance with proposals circulated to cabinet members * No sub-committees of cabinet * No Ministers trained	* 15% compliance with country target for circulating cabinet proposals * Committee structures approved * 10% of Ministers attend training and rate it good or excellent	* 30% compliance with country target for circulating cabinet proposals * Committees interrogate proposals * 30% of cabinet agenda items considered by committees * 25% of Ministers attended	* 50% compliance with country target for circulating cabinet proposals * Committees functioning without external support * 40% of items considered by committees * 40% of Ministers attended	

		Achieved					
		Source					
		Quarterly reports; Cabinet Secretariat monitoring tools and data; training records; discussions with beneficiaries					
	Output Indicator 1.3		**Baseline**	**Milestone 1**	**Milestone 2**	**Target (date)**	
	Line ministries are better able to develop evidence-informed proposals, as measured by: - Network of Cabinet Focal Persons (CFPs) in Ministries established and functioning - Percentage of Ministries with trained CFPs (cumulative) - Number of training days delivered to CFPs	**Planned**	*No cabinet focal persons (CFPs) in Sierra Leone and Liberia * 7.6% of ministries with trained CFPs * No training	* CFPS nominated * Purpose of CFPS agreed by Ministers * Training strategies agreed	* CPFs in place and supported * 60% of ministries with a trained CFP * 1,000 person training days delivered	* CFP network self-sufficient * 75% of ministries with trained CPFs * 2,500 person training days	
		Achieved					
		Source					
		Quarterly reports; training records; discussions with beneficiaries					
IMPACT WEIGHTING (%)	**Output Indicator 1.4**		**Baseline**	**Milestone 1**	**Milestone 2**	**Target (date)**	
20%	Project guidelines, advice and training materials are shared effectively with others, particularly African Cabinet Secretaries, as measured by: - Participants in African Cabinet Development (ACD) network who assess international activities as 'good' or 'excellent' - Number of high-level workshops held - ACD Evidence-based Policy Toolkit is developed and disseminated - Number of media articles covering programme activities (cumulative)	**Planned**	* No materials	* 35 participants in ACD network who rate as good or excellent * 1 high-level workshop * proto-type toolkit * 9 articles on programme activities, of which 6 are in beneficiary countries	* 70 (culm.) participants in ACD * 2 high-level workshops * toolkit developed * 18 news articles, 12 in beneficiary countries	110 (culm.) participants * 3 high-level workshops * toolkit upgraded and subject to at least 40 requests * 25 news articles (18 in beneficiary countries)	
		Achieved					
		Source					**RISK RATING**
		ACD reports and feedback; newspaper or electronic articles					High, given instable operating environment (South Sudan) and high levels of political buy-in required.

OUTPUT 2	Output Indicator 2.1		Baseline	Milestone 1	Milestone 2	Target (date)	Assumptions
Greater use of evidence to inform policy decisions in India and Pakistan (see nested logframe 2)	High quality assessment report completed, as measured by: - Survey and data instruments developed - Data collected and analysed	**Planned**	No available assessment	* Assessment instrument draft, piloted and refined (February 2014) * At least 250 observations * Analysis of training needs of initial training cohorts completed	* Instruments rolled out and further refined * Additional 150 observations * Preliminary data analysis from other instruments	* Instruments made public * End data set of 500 observations * End-line data analysed and assessment report complete	1) Partner organisations willingly participate in data collection and training activities 2) That training participants return to an environment that allows them to use their learning 3) Increased capacity to understand and produce evidence-based policy proposals leads to increased number of evidence-based policy proposals.
		Achieved					
		Source					
		Assessment instrument developed for the project					
	Output Indicator 2.2		**Baseline**	**Milestone 1**	**Milestone 2**	**Target (date)**	
	Curriculum materials developed, as measured by: - Number of online modules developed and tested - Number of civil servants trained in full set of modules - Level of proficiency in technical skills - Attitudes towards use of evidence in decision making	**Planned**	No materials developed for the country contexts	* 2 modules developed (1 day training) * At least 80 civil servants * Specific measures for learning rubric developed to assess changes in trainees' technical skills and attitudes * Baseline data collected among initial training cohorts in all focus countries	* At least 120 civil servants * 6-8 modules developed (3 to 4 training days)	* At least 300 civil servants	
		Achieved					
		Source					
		Course materials developed					
	Output Indicator 2.3		**Baseline**	**Milestone 1**	**Milestone 2**	**Target (date)**	

	Pilot projects successfully implemented, as measured by: - Number of demonstration and pilot projects selected for funding and completed due diligence process (cumulative) - Number of case studies developed, based on demonstration / pilot projects	**Planned**	No pilot projects	* At least 5 demonstration projects	* 3 pilot projects selected	* 6 pilot projects selected * 6+ case studies	
		Achieved					
		Source					
		Data and reporting on demonstration projects and pilot projects					
	Output Indicator 2.4		**Baseline**	**Milestone 1**	**Milestone 2**	**Target (date)**	
	Policy dialogues held, as measured by: - Number of policy workshops held - Number of people attending workshops, including number of female presenters (cumulative) - Number of policy dialogue reports	**Planned**	None	* 2 policy workshops held by December 2014 * 60 attendees to workshops with 4 female presenters by December 2014 * 2 policy dialogue reports by July 2014	* 4 policy workshops held by December 2015 * 120 attendees to workshops, with 8 female presenters by December 2015 * 4 policy dialogue reports by July 2015	* 6 policy workshops held by July 2016 * 180 attendees to workshops with 12 female presenters by July 2016 * 6 policy dialogue reports by July 2016	
		Achieved					
IMPACT WEIGHTING (%)		**Source**					**RISK RATING**
15%		Records of policy dialogue workshops through quarterly reports and beneficiary feedback					Medium
INPUTS (£)	**DFID (£)**		**Govt (£)**	**Other (£)**	**Total (£)**	**DFID SHARE (%)**	
INPUTS (HR)	**DFID (FTEs)**						

OUTPUT 3	Output Indicator 3.1		Baseline	Milestone 1	Milestone 2	Target (date)	Assumptions
Improving the skills, systems and environments to use evidence within the governments and parliaments in Ghana, South Africa and Zimbabwe (see nested logframe 3)	Policymaking staff from selected countries have improved skills for and understanding of Evidence-Informed Policymaking (EPIM), as measured by: - Tailored course for Civil Service Training College (CSTC) in Ghana developed and implemented - Number of public institutions participating in training in Zimbabwe - Changes to South African Government processes to increase the use of evidence - Support provided to Ghanaian and South African parliaments - Number of policy dialogues and knowledge cafes held in Zimbabwe	**Planned**	*No existing courses that support the skills for EPIM *Facilitators do not receive pedagogy training or refresher courses on a regular basis *Facilitators have not worked on courses for EIPM in the past	*MOUs signed with CSTC in Ghana and departments (where appropriate) *EIPM course content developed or adapted from existing *Trainers in civil service colleges identified	*Trainers at the CSTC receive pedagogy and EIPM training *EIPM course/modules trialled with 1 cohort	*EIPM course/modules trialled with 2 cohorts and adopted by CSTC in Ghana	1) Elections in three target countries and other external events do not result in a change of political or high-level support 2) That participants on the course return to an environment that allows them to use their new skills 3) That there is sufficient public appetite for discussions around EIPM in Zimbabwe 4) That consortium partners have sufficient skills to deliver project activities effectively
			Needs assessment demonstrates: - Lack of awareness of benefits of EIPM - Demand from policymakers for support for their staff - Lack of expertise & skills to use & manage research - Poor communication of research	* Agreement reached with 3 institutions in Zimbabwe * EIPM course content developed	* EIPM course content trialled with 3 cohorts * EIPM champions identified (at least 2 per institution) * Mentoring programme designed	* 6 EIPM champions mentored in how to improve use of evidence in their departments * EIPM course delivered to 3 institutions in Zimbabwe	

			Current state of evidence use in South African ministries to be determined through baseline survey	* Collaborating departments selected, with project engagement starting in at least one department * Improved capacity of Human and Social Research Council (HSRC) in South Africa to facilitate processes	* Approaches to improve management of the evidence base developed and reviewed * Second government department identified * HSRC share process of supporting govt departments with other consortium partners	*Lesson learning documents for work with government departments articulating the benefits of using evidence management approaches/tools *HSRC capacity developed to be able to handle future demand
			Baseline to be set following review of parliamentary research structure in year 2 (Ghana) and engagement with portfolio committee (South Africa)	Familiarisation meetings with parliament and parliamentary research directorate in Ghana	* Review of parliamentary research structure in Ghana * EIPM awareness for MPs in Ghana * Parliamentary staff trial EIPM course in Ghana * Engagement with relevant portfolio committee to explore how to scrutinise the use of evidence in the policymaking process in SA	*Increased capacity of staff to use evidence + further demand for capacity building from GH parliament *Parliamentary committees engage to explore how to better scrutinise policy and the use of evidence in SA

		Zimbabwe: 2 knowledge cafes in 2012	1 Policy dialogue and 1 knowledge café in Zimbabwe	3 Policy dialogues and 1 knowledge café in Zimbabwe	*6 policy dialogues and 3 knowledge cafes, with 50% focused on issues that disproportionately impact women. *Media coverage of policy dialogues *Café and dialogues routinely attended by a wide range of stakeholders
	Achieved				
	Source				
	Annual project reports; end of project evaluation; civil service school course list; formal and informal media reports				
Output Indicator 3.2		**Baseline**	**Milestone 1**	**Milestone 2**	**Target (date)**
Number of case studies and other communication outputs from the small grants programme and project consortium on building capacity for research use.	**Planned**	0	4 small grant projects identified and funded	3 case studies published from small grant projects 8 projects identified and funded since start of programme	6 case studies published (cumulative)
		N/A	3 communication outputs	6 communication outputs (cumulative)	*12 communication outputs (cumulative) * Consortium symposium and learning conference held
	Achieved				
	Source				
	Blogs; case studies; annual reports				
Output Indicator 3.3		**Baseline**	**Milestone 1**	**Milestone 2**	**Target (date)**

	Consortium partners are better able to deliver capacity-building activities, as measured by: - Improvements in partners' systems, processes and/or staff kill levels - Demand from others for support (outside of project beneficiaries)	**Planned**	Organisational assessment demonstrates: - Partners have limited capacity (skills and experience) implementing M&E plans and strategies (Ghana and Zimbabwe) - Partners have some capacity (skills and experience) using project & financial management systems - Partners have sufficient pedagogical skills, capacity and knowledge of EIPM	* All partners have a M&E plan in place * All consortium staff who will be directly responsible for delivering training refresh their training skills	* Partners use collaborative project management tools * Partners use M&E tools and templates to collect data	* Partners improve their capacity to develop and implement an M&E plan * Partners show clear improvement in financial and project management * Partners show improvement in their pedagogical skills and knowledge on EIPM	
IMPACT WEIGHTING (%) 20%			* Partners have limited capacity (skills and experience) designing and implementing communication plans and strategies (Ghana and Zimbabwe) * Partners have limited capacity (skills and experience) to develop and use some communications tools	* South Africa: Identification of appropriate personnel in HSRC and training by ODI in application of demand-side toolkit * Communications strategy work plan developed	* HSRC team leads on application of the toolkit in at least one Ministry * Zimbabwe partner identifies champions in key ministries for mentoring support * Ghana partner works with parliamentary resource department to	* Partners show capacity to develop and implement a communication plan * Request to support capacity building from at least one non-project department or committee in all consortium partner countries	**Risk rating** Medium: Elections are expected in all partner countries. The range (types, location and organisations) of consortium activities is spread out which should go some way to mitigating this risk. The potential impact of the risk in a specific area is high e.g. elections may

					develop training plan		impact on the feasibility of policy dialogues in Zimbabwe or change the priorities of the civil service in any one country
		Achieved					
		Source					
		Consortium inception phase capacity assessment report; members post-consortium work plan; end of project evaluation					

IMPACT WEIGHTING (%)	**DFID (£)**	**Govt (£)**	**Other (£)**	**Total (£)**	**DFID SHARE (%)**
20%					
INPUTS (HR)	**DFID (FTEs)**				

OUTPUT 4	**Output Indicator 4.1**		**Baseline**	**Milestone 1**	**Milestone 2**	**Target (date)**	**Assumptions**
Civil servants in South Africa and Malawi have improved capacity and support to use evidence to inform policy (see nested logframe 4)	Project governance and the Africa Evidence Network, as measured by: - Number of needs assessments and partnerships with public policy and delivery partners - Core resources on capacity building developed, including new mentoring and secondment functions	**Planned**	No governance arrangements in place	* Landscape reviews and needs assessments completed * Existing resources (training materials) on capacity building and mentoring systems published * 150 members of Africa Evidence Network, participation at colloquium & use of website	To be agreed once baseline is set: number of secondments for South Africa and Malawi To be agreed once baseline is set: number of partnerships with institutions to deliver capacity-building activities	To be agreed once baseline is set	1) That mentored personnel at government levels will go on to mentor others 2) Sufficient senior-level buy-in to gain traction for reforms with ministries. 3) That participants return to an environment that allows them to use their new skills, following training/mentoring etc.
		Achieved					

		Source				
		Data collected from landscape reviews, needs assessments and other fieldwork.				
	Output Indicator 4.2		**Baseline**	**Milestone 1**	**Milestone 2**	**Target (date)**
	Project raises awareness of evidence-informed policymaking and enhancing capacity in research use among civil servants, as measured by: - Number of training workshop places - Examples of increased use of evidence in policy documents - Improved ability of workshop participants to assessment and synthesise research	**Planned**		* Pilot workshops delivered in South Africa for 40 people (min 30% female) and learning integrated into year 2 plans * At least 1 policy paper reviewed or developed using BCURE support using research evidence in conjunction with partner agency * Engagement with senior personnel	To be agreed once baseline is set: percentage able to assess and synthesise research	To be agreed once baseline is set: number of examples of use of evidence in policy documents
		Achieved				
		Source				
		Pre- and post-training surveys, Follow-up surveys, Stakeholder interviews, Policy documents				
IMPACT WEIGHTING (%)	**Output Indicator 4.3**		**Baseline**	**Milestone 1**	**Milestone 2**	**Target (date)**
15%	Further support mechanisms established that enhance the application of learning among civil servants, as measured by: - Number of male and female civil servants mentored - Number of male and female civil servants seconded on experiential work placements - Case studies of good practice developed and shared	**Planned**	* 0 mentoring relationships * 0 secondments * Invited to present at review of the 2-year national policy-research-nexus meeting (4/14); Invited to contribute to annual reflection	Five pilot mentoring relationships complete Two secondments complete Invitations to one key national-level meeting per quarter; membership of	To be agreed during inception phase	* 20 women and 20 men mentored * Other targets to be agreed during the inception phase

			meeting of National Evaluation Strategy (4/14); Invited to strategic review of PSPPD (5/14).	one strategic steering group			
		Achieved					
		Source					**RISK RATING**
		Mentorship reports; follow-up surveys; email records					Medium
INPUTS (£)	**DFID (£)**	**Govt (£)**	**Other (£)**	**Total (£)**	**DFID SHARE (%)**		
INPUTS (HR)	**DFID (FTEs)**						

OUTPUT 5	Output Indicator 5.1		Baseline	Milestone 1	Milestone 2	Target (date)	Assumption
Improved use of evidence for health policy in Kenya and Malawi (see nested logframe 5)	Optimised institutional leadership and capacity to enhance evidence use: - Number of leaders in MoH and parliament and evidence champions engaged to advocate for their active role in addressing bottlenecks to evidence use - Number of research evidence use sessions held at high-level symposia/meetings in MoH and parliament and health research conference/seminar - Number of sessions held at existing regional fora to promote research prioritisation - Number of activities linking policy institutions, research institutions, policymakers and researchers	**Planned**	* 0 * 0 * 0 * 0	* 22 leaders in MoH engaged (9 & 13 in Kenya & Malawi, respectively); 18 leaders in parliament respectively (11 & 7 in Kenya and Malawi, respectively); recruited 20 evidence champions (12 & 15 in Kenya and Malawi, respectively) * 1 research evidence meeting held in Kenya; 0 held in Malawi * 1 sessions held at Directors' Joint Consultative Committee (DJCC) * 4 policy science cafes held (3 in Kenya and 1 in Malawi)	* 20 leaders in MoH engaged (10 in each country); 14 leaders in parliament engaged (7 in each country); 20 evidence champions engaged (10 in each country) * 2 meetings held (1 health research conference in each country) * 2 sessions held (1 session at DJCC & 1 session with Health Ministers) * 4 policy science café (2 in each country); at least 80% participants giving positive assessment of the policy science cafes	* 20 leaders in MoH engaged (10 in each country); 14 leaders in parliament engaged (7 in each country); 20 evidence champions engaged (10 in each country) * 4 meetings held (2 health research conference in each country) *5 sessions held (2 sessions with DJCC & 2 sessions with Health Ministers and 1 Best Practices forum) * 12 policy science cafes held (7 in Kenya and 5 in Malawi); at least 80% participants giving positive assessment of the policy science cafes	1) Enhanced evidence use in decision making will result in an increase in evidence-informed health policies 2) Increased capacity of mid-level policymakers to use research evidence/data in decision making will result in an increase in evidence-informed health policies 3) Effectively managing and coordinating the programme will result in its effectiveness in improving the capacity of policymakers to use or consider research evidence in their decision making processes
		Achieved					
		Source					
		To be agreed in inception phase					
	Output Indicator 5.2		**Baseline**	**Milestone 1**	**Milestone 2**	**Target (date)**	

	Enhanced capacity of mid-level policymakers in MoH and Parliament in use of research evidence, as measured by: - Number of mid-level policymakers from MoH and parliament trained in use of research evidence - % trainees reporting that the training workshop improved their knowledge and skills immediately after the training workshop and 1 year after workshop - Number of parliamentary clerks participating in UK POST internship programme	**Planned**	* 0 * 0 * 0	* 40 mid-level policymakers trained (20 in each country consisting 15 from the MoH and 5 from parliament) * 80% * 2 parliamentary clerks/research officers (1 in each country); 2 briefs generated by interns; 2 workshops facilitated by interns	* 30 mid-level policymakers trained in both Kenya and Malawi * 80% * 2 parliamentary clerks/research officers (1 in each country); 2 briefs generated by interns; 2 workshops facilitated by interns	* 40 mid-level policymakers trained (20 in each country consisting 15 from the MoH and 5 from parliament) * 80% *4 parliamentary clerks/research officers (1 in each country); 4 briefs generated by interns; 4 workshops facilitated by interns
		Achieved				
		Source				
		To be agreed in inception phase				
	Output Indicator 5.3		**Baseline**	**Milestone 1**	**Milestone 2**	**Target (date)**
	Effective Programme Management and Coordination: - Number of Consortium planning meetings and DFID BCURE Partners Planning meetings held to assess progress and plan for the coming year - Number of meetings of the Programme Advisory Committee (PAC) and mid-term review of the programme in each country - Introduction of a robust financial and programme management systems	**Planned**	*0 *0 *0	*2 meetings held (1 SECURE Health Program Partners Planning meeting & 1 DFID BCURE meeting); record of programme enhancements as a result of attendance of BCURE meeting. * 6 meetings held (2 meetings for PAC (1 in each country); 4 Steering Committee meetings) *Financial and	*2 meetings held (1 SECURE Health Program Partners Planning meeting & 1 DFID BCURE meeting); record of programme enhancements as a result of attendance of BCURE meeting. * 6 meetings held (2 meetings for PAC (1 in each country); 4 Steering	*6 meetings held (3 SECURE Health Program Partners Planning meeting & 3 DFID BCURE meeting) * 19 meetings held (3 in each country for PAC and 12 Steering Committee meetings; 1 mid-term review meeting) * Efficient financial and programme management systems in place

				programme management systems procured and operationalised	Committee meetings; 1 mid-term review meeting) * Financial and programme management systems monitored and evaluated		
		Achieved					
IMPACT WEIGHTING (%)		**Source**					**RISK RATING**
15%		To be agreed in inception phase					Medium
INPUTS (£)	**DFID (£)**		**Govt (£)**	**Other (£)**	**Total (£)**	**DFID SHARE (%)**	
INPUTS (HR)	**DFID (FTEs)**						

OUTPUT 6	**Output Indicator 6.1**		**Baseline**	**Milestone 1**	**Milestone 2**	**Target (date)**	**Assumption**
Improved use of evidence in government decision making in Bangladesh (see nested logframe 6)	Government Policy formulation procedures are evidence based, as measured by: - Policy development procedures produced centrally which mandate the use of evidence - Methodologies, guidelines and templates to support the evidence-based policy development procedures are produced	**Planned**	Current procedures do not mandate this and documents do not support evidence-based approach	To be confirmed during inception phase	To be confirmed during inception phase	Target ministries adopted procedures and guidance	1) There is sufficient senior-level buy-in to gain traction with Ministries for training 2) Local research organisations are able and willing to work with government ministries 3) Senior-level buy-in from Cabinet Secretary and
		Achieved					
		Source					
		To be agreed in inception phase					
	Output Indicator 6.2		**Baseline**	**Milestone 1**	**Milestone 2**	**Target (date)**	

	Improved ability in line ministries to follow evidence-based policy formulation process, as measured by: - Number of policy proposals produced in target line ministries which incorporate evidence in their development - Scores of Line Ministry officials on pre- and post-training tests for training on ex-ante assessments and evidence literacy	**Planned**	0 officials achieving a 25% increase	Milestones on policy proposals to be agreed during inception phase 30 officials achieve 25% increase	Milestones on policy proposals to be agreed during inception phase 60 officials achieve 25% increase	Milestones on policy proposals to be agreed during inception phase 90 officials achieve 25% improvement on their capacity to use evidence	Ministers to agree and implement government-wide processes and systems to increase use of evidence
		Achieved					
		Source					
		To be agreed in inception phase					
	Output Indicator 6.3		**Baseline**	**Milestone 1**	**Milestone 2**	**Target (date)**	
	Greater collaboration between line ministries and local research providers, as measured by: - Number of policy proposals in target line ministries which featured evidence or inputs from local research providers - MoUs signed between target line ministries and local research providers	**Planned**	To be confirmed - Based on number of proposals in target line ministries that include evidence or inputs from local researchers	Baseline +5 MOU milestones to be agreed during inception phase	Baseline +8 MOU milestones to be agreed during inception phase	Baseline +10 MOU milestones to be agreed during inception phase	
		Achieved					
		Source					
		To be agreed in inception phase					
IMPACT WEIGHTING (%)	**Output indictor 6.4**	**Planned**					
15%	Research is made available on factors which influence the uptake of evidence-based policymaking within each of the line ministries, as measured by: - Assessment frameworks are developed for each target line ministry - Assessment frameworks are applied at mid-point and end point of support to target line ministry	**Planned**	To be confirmed - based on assessment frameworks which will be developed for each ministry	3 frameworks	6 frameworks developed	6 frameworks and assessments undertaken	
							RISK RATING
		To be agreed in inception phase					Medium

2 BCURE management and learning

The BCURE programme was managed through an overarching logical framework that aggregates the component programmes (see Annex 1). The individual BCURE projects each had their own logframes and programme managers (from DFID's Evidence into Action team). The portfolio was not expected to work as a 'sum of the parts' programme. However, all the implementing partners and their DFID programme managers shared learning from their programmes on strategies and approaches (e.g. training curricula) and collaborated if appropriate.

Programme teams participated in an annual learning event facilitated by DFID, supported by an online communications platform, managed by DFID.[3] The BCURE evaluation also fed into the cross-programme learning by sharing findings at the learning events. DFID staff led and facilitate the internal learning and knowledge exchange aspects of the programme. The evaluation team led on communicating the evaluation findings with a wider audience to promote uptake and use.

Key audiences for the evaluation

The evidence base on capacity development for EIPM is small, largely derived from the health field, and weighted towards studies examining the impact of training on individual capacity. There are significant evidence gaps around the role of interpersonal and organisational interventions in promoting change, and regarding the influence of EIPM capacity development on policy change and improved quality of policy development processes. There is a particular lack of evidence on capacity development for EIPM in developing countries. Operational insights into how to design and implement this type of intervention in developing country contexts are also lacking.

To strengthen this evidence base, the BCURE evaluation provides robust evidence on how and why different approaches to capacity building for EIPM work, for whom and in which contexts, in developing countries. These lessons are intended to be directly applicable to the commissioning, design, implementation and adaptation of EIPM capacity-building programmes in developing countries to improve results.

Therefore, the intended users of the synthesis report are, in the first instance, BCURE's managing team at DFID's Research and Evidence Division and the BCURE partners responsible for delivering BCURE programmes, to inform improvements within the current portfolio of programmes.

The findings are also intended to be of use to a wider audience of donors, funders, commissioners and implementers who are considering future EIPM capacity development programmes. These evaluation users may be in numerous fields, such as governance, public management and administration, and research and evidence utilisation. For these audiences, the evaluation findings provide evidence on:

1. How and why different interventions lead to change, and contextual factors that affect outcomes.
2. How interventions can be combined in multi-level capacity development strategies.
3. How and why capacity development interventions can contribute to organisational and institutional shifts to embed EIPM behaviours and systems, ultimately enhancing policy development processes.

An evaluation communications framework was developed to facilitate the contribution of the evaluation to the wider evidence base on EIPM, and a range of communication activities have been conducted and will continue following the publication of the final evaluation. Annex 10 provides further information.

[3] See https://bcureglobal.wordpress.com/

3 Evaluation design and methodology

This section presents the full evaluation design and methodology, expanding on the summarised version contained in the main report.

3.1 Evaluation questions

The BCURE evaluation addresses two overarching evaluation questions (EQs). These are based on the questions posed in the Terms of Reference (Annex 1), revised in the inception phase following discussions with DFID.

1. How effective are the BCURE projects in achieving their stated outcome of increasing the use of evidence in public sector decision making, and influencing longer-term changes in policy quality?
2. How and why does capacity building for evidence use work/not work, for whom, to what extent, in what respects and in what circumstances?

The initial evaluation framework identified ten evaluation questions underlying the two overarching EQs, which were designed to test different parts of the common theory of change (CToC). This proved to be unwieldy, and the framework was streamlined for Stage 2. It was decided to focus on five questions, built around four domains of capacity change (individual, interpersonal, organisational and institutional) within our programme theory, as follows:

Stage 2 Evaluation questions

EQ 1. How and why did BCURE contribute to individual-level change? 1.1 What outcomes were achieved? 1.2 How did the interventions lead to outcomes? *(Testing Stage 1 CIMOs 1, 2, 3, 4, 5)* 1.3 How sustainable were the outcomes? 1.4 What was BCURE's contribution to the outcomes?
EQ 2. How and why did BCURE contribute to interpersonal-level change? 2.1 What outcomes were achieved? 2.2 How did the interventions lead to outcomes? *(Testing Stage 1 CIMOs 7, 8)* 2.3 How sustainable were the outcomes? 2.4 What was BCURE's contribution to the outcomes?
EQ 3. How and why did BCURE contribute to organisational-level change? 3.1 What outcomes were achieved? 3.2 How did the interventions lead to outcomes? *(Testing Stage 1 CIMOs 6, 9, 10, 11, 12 13, 14)* 3.3 How sustainable were the outcomes? 3.4 What was BCURE's contribution to the outcomes?
EQ 4. How and why did BCURE contribute to institutional-/system-level change? 4.1 What outcomes were achieved? 4.2 How did the interventions lead to outcomes? *(No CIMOs identified in Stage 1; to be developed at Stage 2)* 4.3 How sustainable were the outcomes? 4.4 What was BCURE's contribution to the outcomes?
EQ 5. How and why did BCURE (and similar EIPM capacity-building interventions) contribute to changes in policy quality? 5.1 What outcomes were achieved? 5.2 How did the interventions lead to outcomes? (No CIMOs identified in Stage 1; to be developed at Stage 2) 5.3 How sustainable were the outcomes? 5.4 What was BCURE's contribution to the outcomes?

Several revisions to the evaluation design were undertaken at Stage 3 in response to comments from the evaluation Steering Committee and DFID independent quality review at Stage 2. One key piece of feedback was that framing the evaluation around the four levels of change potentially sacrificed clarity and the ability to understand and describe findings in a holistic, case-based way. There were also concerns that the scope of the evaluation was quite broad, involving exploration of a wide range of emerging outcomes in six BCURE case study countries. This produced broad evidence of outcomes but did not provide the depth of evidence to draw definitive conclusions. At Stage 3, it was therefore decided to focus on a smaller number of 'priority outcomes' rather than investigate all of the anticipated outcomes across the BCURE projects.

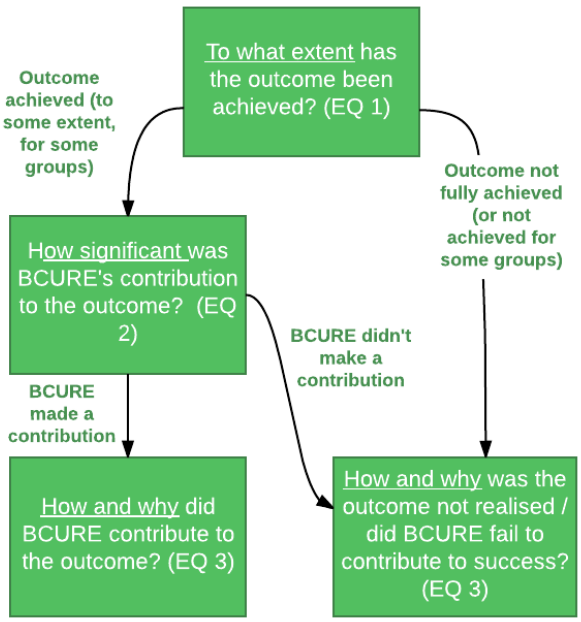
In line with this, the Steering Committee agreed to further revise the EQs. Rather than linking them to the different levels of change, they were linked to priority outcomes identified within the country programme theories.

Stage 3 Evaluation questions

EQ 1. To what extent have priority outcomes been realised and for whom, and how sustainable are they? Have the theorised changes happened? How far have these changes occurred across different sub-groups and organisations etc., reflecting on gender and equity issues? How sustainable are the changes?
EQ 2. How significant was BCURE's contribution to priority outcomes, alongside the contribution of non-BCURE factors? What is the evidence that BCURE contributed to causing the observed changes, and what is the evidence that non-BCURE factors contributed? What is the relative importance of BCURE and non-BCURE factors in explaining the observed changes?
EQ 3. How and why did BCURE contribute or fail to contribute to priority outcomes? Through which mechanisms, enabled by which features of the intervention and features of the (individual, interpersonal, organisational and institutional) context, did BCURE contribute to the observed changes?

To answer the three EQs, the Stage 3 evaluation gathered and analysed evidence from various sources against country-level theories of change, to first judge the extent to which an expected outcome had emerged (EQ 1), then establish the extent to which BCURE contributed to this outcome (EQ 2), and finally determine how, why, for whom, and in what circumstances the outcome had or had not happened (EQ 3). Figure 1 depicts the logical flow of the evaluation questions, which was used to structure the approach to data collection and analysis. As agreed with the evaluation Steering Committee, the evaluation questions were framed around case-specific priority outcomes and thus were answered at the level of the internal country case study reports. This overview report provides summary comparative judgements across the cases in relation to the EQs, but its purpose is not to answer the questions at a portfolio level.

Figure 1. Logical flow of the EQs



3.2 Approach to answering the evaluation questions

The BCURE interventions work in complex government contexts, with myriad contextual conditions influencing potential outcomes. These included diverse historical institutional trajectories; varied political and economic conditions, government systems and organisational cultures; and a wide range of participant characteristics (individuals' identities, gender and ethnicities). Quasi-experimental and counterfactual approaches are unsuited to evaluating this type of programme, as there is no possibility of establishing a control group or comparator (Stern et al., 2012). In addition, BCURE was likely to be just one of a number of factors influencing change in complex government systems, giving rise to the 'attribution problem' – the challenge of attributing a particular change to a particular programme when other factors are also contributing (Wimbush et al., 2012).

In order to address these challenges and answer the evaluation questions, the evaluation adopted a **realist evaluation** approach, drawing on elements of **contribution analysis** and taking a **political economy lens**.

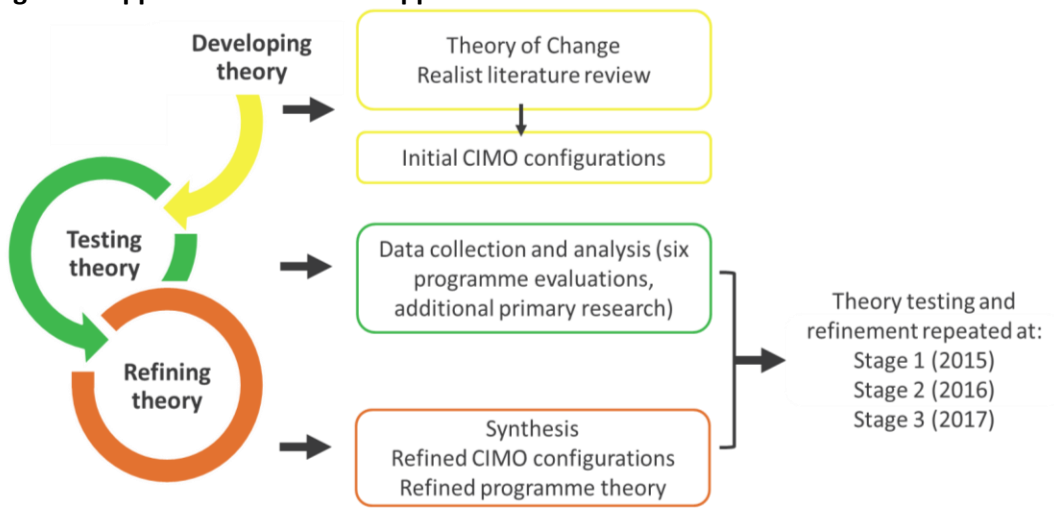
Realist evaluation

A realist approach was selected because the primary aim of the evaluation was to strengthen the evidence base on how capacity building can promote EIPM, to inform decisions within and beyond DFID about whether to fund and how to design this type of programme in future. DFID was interested in understanding not just *whether* BCURE worked but also *how and why* capacity building can contribute to increased use of evidence in policymaking in the very different contexts in which the programme is operating (EQ 3). Realist evaluation works through opening up the 'black box' between interventions and outcomes, through developing and testing *programme theory* (an explanation of how, why, and in what contexts interventions lead to particular outcomes – see Box 1).

Programme theory consists of linked sets of hypotheses about the *mechanisms* that cause an intervention to work or not work in particular *contexts*, to lead to specific *outcomes*. These hypotheses are known as 'context–mechanism–outcome' or CMO configurations (see Box 1) – the core analytical units of realist evaluation (Pawson and Tilley, 1997; Wong et al., 2013). The evaluation team decided to incorporate *features of the intervention* as an additional element to our CMO configurations, in order to separate out features that are inherent in or under the control of the programme (such as training design or length) from contextual factors that are not (such as professional incentives to participate in training) when considering what might 'spark' a particular mechanism. This gives us the formulation C+I+M=O (CIMOs), used throughout this report.

Realist evaluation encompasses three broad stages: developing theory, testing theory and refining theory. These are iterative rather than linear; theory is developed, tested, refined and tested again as knowledge accumulates. Figure 2 provides an overview of the evaluation design.

Figure 2. Application of a realist approach in the BCURE evaluation



Box 1: Context, mechanism, outcome and programme theory

Mechanisms are the causal forces, powers, processes or interactions that generate change within an intervention – including the choices, reasoning and decisions people make as a result of the resources the programme provides. An intervention such as a training course is not a mechanism. The mechanism is the 'thing' that explains *why* training changes behaviour (or does not) in a particular setting.

Mechanisms are triggered only in certain **contexts**. Contextual factors may include *individual* characteristics that affect how people respond to opportunities (e.g. gender, ethnicity, education); *interpersonal* factors that affect trust and buy-in (relationships between stakeholders and programme implementers); *institutional* factors (the rules, norms and culture of the organisation in which the intervention is implemented); and *infrastructural* factors – the wider social, economic, political and cultural setting of the programme (Pawson and Tilley, 2004).

Outcomes refer to intended and unintended short-, medium- and long-term changes resulting from an intervention.

A **CMO configuration** is a theory or hypothesis about how a particular mechanism works in a specific context to lead to an outcome. They can usually be read as sentences – for example, 'Where training content is directly relevant to a person's day job (C), providing information about how evidence can improve policymaking can spark an "eye-opener" in which trainees recognise how evidence can add value (M), leading to increased use of evidence in their day-to-day work (O)'.

A **realist programme theory** explains '(some of) how and why, in the 'real world', a programme 'works', for whom, to what extent and in which contexts' (Wong et al., 2016). A realist programme theory is a variation on a ToC, which explicitly spells out the causal links between outcomes as CMO configurations. 'Assumptions' in a ToC are embedded as *theories to be tested* in the CMOs as contextual factors and/or conditions necessary for mechanisms to fire. Some ToC approaches also include 'risks to assumptions' – that is, factors that will prevent the assumptions from holding true. Again, realist programme theory integrates this into the CMO testing, by explaining the contextual or intervention factors that block mechanisms from operating.

Source: Pawson and Tilley, 1997; Westhorp, 2014; Punton et al., 2016b

The first iteration of the BCURE theory drew on the evaluation team's existing knowledge and professional hunches about the nature of capacity building, and how capacity building can contribute to evidence use in policymaking. This was used to shape the research questions for the BCURE literature review, which identified additional theories in the wider literature about how capacity building can contribute to EIPM. These were used to develop our first iteration of CIMO configurations. Stages 1 and 2 of the evaluation then began to test and refine these CIMOs, contributing to a revised programme theory at each stage. At Stage 3, a prioritised set of theories have been tested and revised for a final time, and are presented in the report. Annex 4 contains a full explanation of how the BCURE theory has evolved over time, and lists the CIMOs tested at Stage 3.

Contribution analysis

In order to answer EQ 2, the Stage 3 evaluation drew on elements of contribution analysis. Contribution analysis is a theory-based evaluation approach that provides a systematic way to arrive at credible causal claims about a programme's contribution to change. It allows a robust assessment of cause and effect when it is not practical to design an experiment to measure the attribution of a particular change to a particular

programme (Mayne, 2012). The six steps of contribution analysis[4] provided a framework to help prioritise outcomes and causal links to investigate during Stage 3, and assess the contribution of the programme alongside the role of other factors, as follows:

- A country-level ToC was developed for each case study, allowing the underlying causal logic to be unpacked.
- Evidence from earlier stages of the evaluation was assembled, in order to assess the strength of the existing contribution story, and identify weaknesses and gaps.
- Priority outcomes and causal links to focus on at Stage 3 were then selected, based on a consideration of their importance to the overall contribution story, and utility and importance to stakeholders (Lemire et al 2012).
- Evidence about the extent of BCURE contribution was then collected through country case studies, including through incorporating questions about contribution in the interview topic guides, and examining other explanations for observed outcomes through the political economy lens.
- The country case study analysis then involved a systematic assessment on the extent of BCURE contribution against the country-level ToCs, described further below.

Political economy lens

The Stage 3 evaluation aimed to incorporate a stronger understanding of how political economy issues affect evidence use in policymaking, in order to unpack non-BCURE drivers of outcomes (EQ 2) and incorporate political economy dimensions into our explanations of why BCURE contributed or failed to contribute to outcomes – i.e. the 'C' in CIMOs (EQ 3). A light touch political economy analysis (PEA) exercise was conducted at both country level (to identify key overarching factors and trends that are shaping and influencing policymaking and evidence use) and sector level depending on the sectors targeted by the BCURE partner, as part of each country case study. This was guided by a framework incorporating a checklist of PEA questions, drawing from various pragmatic PEA tools (Hudson et al, 2016; Poole, 2011; Moncrieffe and Luttrell, (2005). The approach is described in more detail in Annex 3.4 below, and the framework is presented in Annex 5.

3.3 Evaluation components

The evaluation had four main components:

1. **Annual programme evaluations of BCURE-funded projects**, incorporating primary data collection within one country (the 'country case study'), and analysis of monitoring and implementation documents from all country contexts. At Stage 3, the evaluation refocused its resources to conduct four evaluations instead of six. This allowed the team to investigate a smaller number of priority outcomes in more depth.
2. **A realist literature review**, synthesising published papers and grey literature related to capacity building for EIPM.
3. **An impact case study**, consisting of additional primary research on a similar intervention to BCURE that had been running for a longer period and therefore closer to seeing 'impact', in order to provide evidence on how capacity building for EIPM contributes to improvements in policy quality (the ultimate goal of the BCURE programme).
4. **A synthesis of findings**, drawing together insights on how and why capacity building for evidence use works or does not work in different contexts.

[4] These six steps are: setting out the cause-effect issue to be addressed; developing a theory of change; gathering existing evidence on the theory of change; assembling and assessing the contribution story and challenges to it; seeking out additional evidence; revising and strengthening the contribution story (Mayne, 2011).

Data collection and synthesis was repeated each year for three years to enable the evaluation to track programme results over time, and iteratively test and refine our theories about how and why particular outcomes have occurred in different contexts – see Figure 1 above. The four components are described in more detail below.

3.4 Component 1. Programme evaluations and country case studies

3.4.1 Approach

During Stage 1 and 2 of the evaluation, programme evaluation reports were produced for each of the six BCURE projects. At Stage 3 it was agreed with the Steering Committee that the evaluation would conduct four 'country case studies' instead of six programme evaluations, to enable a focus on 'depth' rather than 'breadth'. The reports performed two functions:

- **Providing internal management reports for each project**, which verified outcomes identified by the BCURE programme monitoring data (and identified additional outcomes), captured key lessons and recommendations and generated an assessment on programme effectiveness, value for money, sustainability and programme contribution to change, in order to inform decision making.
- **Collecting data on how and why BCURE projects contributed to different patterns of outcomes**. This data was then fed into the synthesis, in order to identify, test and refine theories about how and why BCURE interventions lead to, or do not lead to, change.

At Stage 3, the programme evaluations were reframed as 'country case studies', and focused primarily on the second function. Each programme evaluation / country case study consisted of an independent review of secondary monitoring data and implementation documents produced by the project team, and primary data collection by the evaluation team within one of the countries targeted by the project. Over the course of the evaluation, 15 programme reports have been produced (five programme evaluations at Stage 1, six at Stage 2, and four country case studies at Stage 3). These are all internal to DFID.

3.4.2 Selection of country case studies

BCURE worked across 12 countries. The evaluation was only able to cover six with available resources. The country case studies were selected during the inception phase using case replication logic (Yin, 2003). Country contexts were grouped into three broad case types based on a typology of anticipated contextual conditions:

1. *Typical*: where the contextual conditions are mixed but could offer some degree of political stability and established institutions to support EIPM.
2. *Challenging*: where the contextual conditions could, according to preconceived assumptions, create difficulties for introducing EIPM.
3. *Favourable*: offering, on first viewing, the most favourable conditions for EIPM – for example a high degree of stability, ordered institutional practices, a good degree of political openness.

Pragmatic considerations of security and access also informed the final selection. Table 2 gives an overview of the countries and the reason for their selection.

Table 2: Country case study selections

BCURE country case study	Case replication logic
Harvard BCURE: Pakistan	The Stage 1 case study focused on India: 'favourable' case (*literal replication*). However, in 2016, activities ceased in India as a result of a refocusing of the UK government's relationship with the country. Pakistan was selected as a replacement as it is the main alternative focus of the Harvard programme. Pakistan is a 'challenging' case (*theoretical replication)*
UJ-BCURE: South Africa Impact case: South Africa	'Favourable' case (*literal replication)*
SECURE Health: Kenya	'Typical' case (*literal and theoretical; both similar and contrasting results possible)*
ACD: Sierra Leone (though Stage 1 evaluation data collection will be difficult)	'Challenging' case (*theoretical replication)*
ECORYS: Bangladesh	'Typical' case *(both similar and contrasting results possible)*
VakaYiko: Zimbabwe	'Challenging' case (*theoretical replication*)

At Stage 3, it was decided to focus on four countries rather than six, in order to allow for a more in-depth investigation. Case studies were selected based on the following considerations:

- The feasibility of accessing data and stakeholders in the context, based on the potential receptiveness of partners given that most projects will have finished by the time the evaluation commences data collection, and also considering other issues that might affect feasibility such as elections.
- Aiming for a balance across African and Asian contexts.
- Aiming to provide insights from different delivery models, i.e. sectoral focus, entry point, number of ministries targeted etc.
- Focusing on DFID countries by spend.

Based on these factors, the following four countries were selected: Bangladesh, Pakistan, Kenya and Zimbabwe.

3.4.3 Methodology for Stage 3 country case studies[5]

The Stage 3 case studies were designed and conducted following six iterative steps:

Step 1. A country-level ToC was developed, drawing on the programme's own ToC, evaluation data from Stages 1 and 2 on outcomes and causal links (CIMOs), and insights from the wider literature.

Step 2. Existing evidence was assembled for each outcome and causal link (CIMO), and gaps and limitations were identified in the existing evidence base, including around political economy dimensions. Priority outcomes and causal theories were identified from this preparatory analysis. In some countries where interventions differed by sector, a sector-specific ToC was developed, which drew on the country-level ToC but reflected specific outcomes and causal pathways.

Step 3. A political economy analysis was conducted to contextualise the ToC within the risks and opportunities that the context posed for EIPM and the programme's desired outcomes.

[5] The methodology for the Stage 1 and 2 programme evaluations is detailed in the Stage 1 and 2 Synthesis Reports and annexes, available from http://www.itad.com/knowledge-and-resources/bcure/

Step 4. Based on the priority outcomes, a purposive sampling framework was then developed to gather secondary data and collect primary data. The sampling process was iterative, developed and revised throughout the data collection process. Data collection involved iteratively triangulating evidence of outcomes, as well as testing and modifying theories about BCURE's contribution to outcomes, the role of other factors, and how and why BCURE contributed or failed to contribute to priority outcomes.

Step 5. A small number of examples of potential policy-level influence were identified by interview respondents, and these were investigated in greater detail through one or two illustrative case studies per country case study.

Step 6. Primary and secondary data was then analysed against evaluation questions to establish the extent of: priority outcomes (EQ 1); BCURE's contribution (or non-contribution) relative to other factors (EQ 2); how and why BCURE contributed or failed to contribute (EQ 3); and examples of policy influence, in order to assemble a summative 'contribution story' for the country case study.

Step 1. Country-level theory of change

Based on the findings from Stage 2, a draft country ToC was developed in the design phase. This aimed to be as specific as possible about the outcomes anticipated by the programme, the critical political economy dynamics affecting the context, and the observed/theorised causal links, to provide a more concrete and contextualised framework for the country case study.

The ToC built on the BCURE programme's own ToC, evidence from Stages 1 and 2 of the evaluation and the broader literature, and consultation with programme staff. It was designed to enable prioritisation of outcomes and causal links for investigation at Stage 3, and systematic investigation of the evaluation questions. The ToC also incorporated our theories (CIMOs) from Stage 2, about how and why certain outcomes were expected to lead to other outcomes in particular contexts, based on evidence collected to date. The ToC was validated with programme teams prior to data collection.

Step 2. Assembling existing evidence for the country-level ToC, and identifying 'priority outcomes'

Drawing on advice about how to test programme theory and insights from contribution analysis, we assembled existing evidence for the country-level ToCs (Funnell and Rogers, 2011; Mayne, 2008; 2012). Data tables were designed to aid this process. Evidence from the Stage 1 and 2 evaluations was collated in relation to each outcome in the country-level ToC, against the three EQs. This allowed us to identify where there was already substantial evidence, and where there were weaknesses and gaps.

Contribution analysis provided a framework to help prioritise outcomes and causal links to investigate during Stage 3. Lemire et al (2012) suggest that prioritisation should be based on a consideration of:

- Fit with purpose of evaluation
- Importance to overall contribution story
- Utility and importance to stakeholders.

Based on these considerations and in consultation with DFID, we identified priority outcomes specific to each country's ToC. We prioritised longer-term outcomes crucial to the overall programme goal of improving the use of evidence in policymaking processes. We also collected evidence against shorter-term outcomes where there were significant evidence gaps that need to be addressed in order to strengthen the overall contribution story for the programme. Finally, we also collected evidence about the political economy dynamics that have shaped BCURE's contribution (or non-contribution).

Step 3. Conducting a political economy analysis

Political Economy Analysis (PEA) is concerned with the interests and incentives of different groups and how they generate policy outcomes; the role that formal institutions and informal norms play in shaping interaction; and the impact of values and ideas on political behaviour and public policy (DFID 2009). The

Stage 3 evaluation aimed to incorporate a stronger understanding of how political economy issues affect evidence use in policymaking. The political economy lens was linked to the revised EQs as follows:

- Helping to investigate *non-BCURE drivers of outcomes*, including the role of interests and incentives, formal institutions and informal norms, and values and ideas (EQ 2).
- Helping to incorporate political economy dimensions into our explanations of why BCURE contributed or failed to contribute to outcomes – i.e. the 'C' in CIMOs (EQ 3).

PEA was conducted at two levels:

1. Country-level, to identify key overarching factors and trends that are shaping and influencing policymaking and evidence use.
2. Sector level: guided by the country theory of change. We defined 1–2 sectors of interest within the country case study contexts, in collaboration with the programme teams and DFID.

A PEA framework was developed, incorporating a checklist of PEA questions drawing from various pragmatic PEA tools (Hudson et al, 2016; Poole, 2011; Moncrieffe and Luttrell, 2005). This was used to structure an initial review conducted by the national consultant prior to data collection, drawing on secondary data sources. Further information was collected through primary interviews with sectoral experts and government stakeholders during the main data collection stage.

Step 4. Developing a purposive sampling framework and collecting data

The priority outcomes guided our sampling and data collection for Stage 3. The aim was to achieve a sufficient degree of confidence about the extent to which priority outcomes had occurred (EQ 1), BCURE's contribution to the outcomes (EQ 2) and how and why BCURE contributed or failed to contribute (EQ 3).

Once priority outcomes were identified for each country case study, we began developing a purposive sampling framework. Our Stage 3 sampling followed four main principles:

1. **Sampling was driven by theory:** In line with our realist evaluation approach, sampling decisions were guided by our theory about the outcomes we expected to observe, and how and why these outcomes are expected to come about – in other words the country-level ToC and associated CIMOs.
2. **Sampling was iterative:** Following from this, sampling was iterative and sampling frameworks flexible, allowing for changes and additions during field work as theories developed and leads were followed. The sample therefore continually evolved throughout the data collection process.
3. **Sampling aided comparison between sub-groups:** A key element of our sampling strategy was comparison between different sub-groups of participants (e.g. more junior and more senior staff, and participants in different roles or ministries), in order to explain differential outcomes.
4. **Sampling sought to maximise triangulation of sources for each theory**: We aimed to triangulate evidence across a range of different stakeholders, through comparing insights from project participants with insights from knowledgeable 'outsiders' (informed by the PEA of who is influential in relation to the outcome), and through accessing secondary documentation where available. Our data sources are detailed below.

We identified stakeholders to interview in two ways:

- **Using previous samples, programme stakeholder lists, monitoring data, and staff recommendations.** During the case study design stage, an initial, incomplete list of interview respondents was identified through reviewing monitoring data and programme documentation (including complete participant lists where available), and conversations with programme staff. Interviews at Stage 3 built on the insights from Stage 1 and 2, and a substantial number of participants were consulted across two or more stages

to allow change to be tracked over time. Decisions about whether to interview the same participants again were based on the four main sampling principles above.

- **Using snowball sampling**. Once fieldwork was under way, interview respondents were asked to identify further individuals who could provide an insight into a particular outcome, or who were non-participants who could help to verify an outcome, for example, colleagues and unit managers. This strategy proved crucial in helping the team to identify knowledgeable non-participants in BCURE interventions, who were unknown to programme staff and therefore difficult to identify up-front.

Data sources

We drew on five types of data, with the aim of triangulating insights for each outcome from as many of the categories below as possible.

1. Monitoring data and other programme documentation, including pre- and post-training course test data, participant feedback on various programme activities, memoranda of understanding with government partners, activity reports, meeting minutes, and case studies written by BCURE partners. This was reviewed first to examine secondary evidence for theories, and to help identify relevant sub-groups of individuals to target for interviews.
2. Interviews and workshops with programme staff. This supplemented the monitoring data, helping understand gaps or areas where greater clarity was needed. They also aimed to explore the team's perceptions on observed changes with different groups, how and why the interventions contributed to change, and blockages to change. It also provided an insight into the areas project staff thought had been more or less successful and how and why, which helped to further develop our theories.
3. Interviews with project participants (individuals directly targeted by the activity / activities which aimed to contribute to the outcome). This generated self-reported insights about the extent to which outcomes had been achieved or not achieved, how and why, for different groups. We considered the possibility of collecting data from a larger number of project participants through a survey, but have rejected this as we felt it would be very difficult to get an acceptable response rate.[6]
4. Interviews with other knowledgeable stakeholders. These were stakeholders who did not participate in BCURE interventions, but who provided insights into (a) outcomes observed and the reasons for these outcomes; and (b) political economy issues that related to priority outcomes.
5. Secondary (non-project) documentation. This encompassed documentation not produced by the programme, which provided insights into outcomes and the reasons for outcomes: for example, government policy or guidance documents. We tried as much as possible to identify and secure potentially relevant documentation, (a) up-front when evidence was assembled; and (b) throughout the data collection phase, using interviews to attempt to secure documents that helped to triangulate insights from respondents.

During data collection, evidence underpinning particular findings was triangulated in three ways:

- Internally, within interviews – claims about change were triangulated through asking for examples and further detail from the respondent.
- Between different interview respondents (different categories of respondent, different individuals within the same department, line managers and line managers, identified through iterative snowball sampling).
- Between primary and secondary data sources.

[6] Several projects have already conducted surveys as part of their M&E, with medium-low response rates, and we felt we are unlikely to get anything better.

The total number of stakeholders consulted for the Stage 3 country case studies is summarised in Table 4. Full lists of respondents are included in the country case study reports. In total, 528 stakeholders were consulted across the six countries over three years.

Table 4: Total number of stakeholders interviewed at Stage 3

	Bangladesh	Kenya	Pakistan	Zimbabwe	Total
BCURE programme staff and direct implementing/consortium partners	6	5	8	6	25
BCURE programme participants	37	24	31	25	117
Non-participants, including government officials, civil society and other external actors	17	20	6	25	68
Total	**60**	**49**	**45**	**56**	**210**

Note: this table does not include interviews with BCURE programme managers

Interview tools

A set of semi-structured interview tools were developed (See Annex 7), designed to be customised to each individual stakeholder. Given the limited time available for interviews, it was necessary to prioritise which elements of the ToC and which CIMOs to test with different stakeholders, especially when respondents were involved in a range of different interventions, theorised to work in different ways.

In order to ensure that we explored outcomes, contribution and CIMOs systematically, we developed an 'Outcomes and CIMO question bank': a set of interview questions that covered the whole ToC. We decided, as part of the sample development and iteratively throughout the data collection, which outcomes and which CIMOs to discuss with which respondents based on relevance and consideration of data gaps. We designed unique interview guides for each respondent that aimed to test the most relevant theories, adding questions from the outcome and CIMO question bank into the generic interview templates.

The sampling spreadsheet was updated after each interview to keep track of which outcomes and which CIMOs had been discussed with which respondents, to ensure that we were testing the country ToC systematically. We used later interviews to corroborate and plug gaps in earlier ones. In addition to the CIMOs we were able to test explicitly, we were also able to infer information relating to our CIMOs from interviews during the analysis. The findings from both explicit and inferred testing were recorded as part of the analysis process.

The country case studies were scheduled to allow the interview guides to be tested by the team leader during the first case study fieldwork. The interview tools and question banks were iteratively revised over the course of the first case study, before subsequent case studies commenced.

Step 5. Embedded case studies on policy-level influence

The ultimate aim of BCURE was to influence the quality of policies in order to improve the lives of poor people. However, it was not feasible for the evaluation to systematically analyse evidence use in all decision-making processes in targeted BCURE organisations and ministries. First, BCURE programmes did not directly target specific policy processes but worked in a broader way to strengthen individual and organisational capacities and processes to enhance evidence use. Second, to focus on the decision-making processes taking place within a ministry would require considerable resources, disproportionate to the anticipated contribution of the programme.

Instead, we sought to (a) systematically catalogue examples of reported policy-level influence through the investigation of lower-level outcomes; and (b) conduct a small number of embedded case studies on a sub-

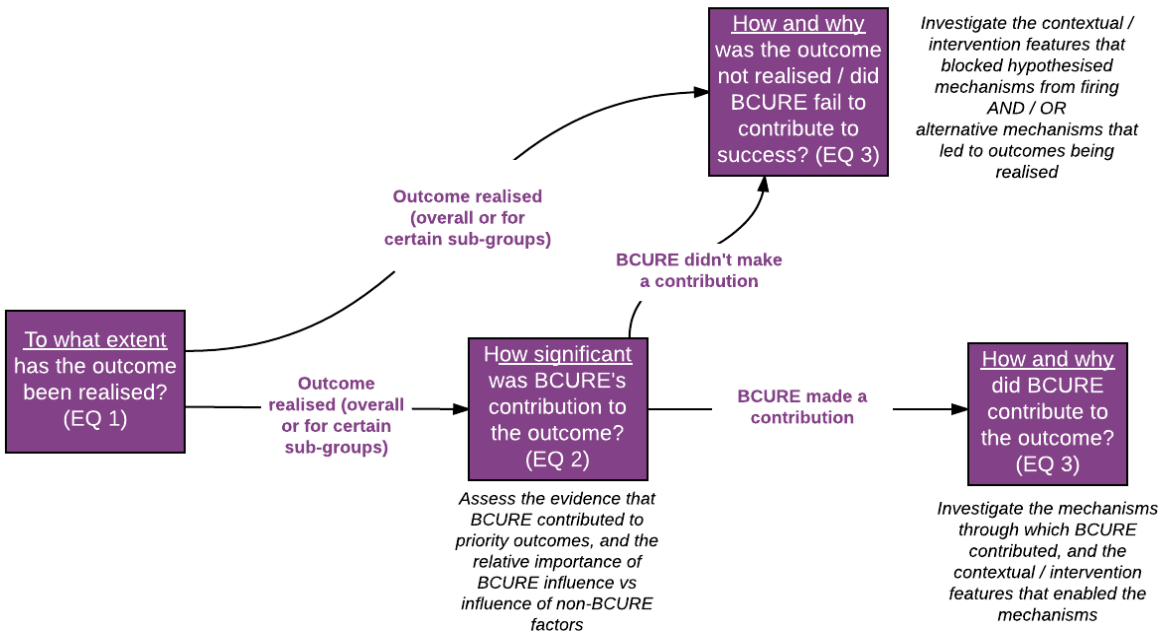
set of these examples. After an initial harvesting of examples of policy influence from respondents, we selected 1-2 examples per case study that were illustrative of an important pathway and appeared credible, triangulated them with supplementary interviews, and analysed them using the EQs.

Step 6. Analysis

Primary data from workshops and interviews was extracted evidence into a Microsoft Excel analysis spreadsheet, as follows:

1. Transcripts were reviewed for insights on the *outcomes* mentioned by respondents, in order to answer EQ 1. Each outcome was entered into a new row in the spreadsheet, in summary form, supported by a relevant quote from the transcript. Where a respondent had also been interviewed in Stage 2, programme leads reviewed the transcript from the previous year, to gain a sense of whether outcomes had been furthered or deepened.
2. Transcripts were the reviewed for insights on the *contribution* of the BCURE programme to the outcome, relative to other factors, including political economy issues, in order to answer EQ 2. This information was entered alongside the outcome in the same row, in summary form, with a supporting quote, as before.
3. For EQ 3, the transcript was then reviewed for the evidence arising from testing the CIMOs to explain how and why the outcome came about: the *mechanisms* respondents felt had contributed to the outcome and the *contextual and intervention factors* respondents felt had enabled (or prevented) the mechanism 'firing'. This process was an interpretive rather than mechanical one, requiring skill and judgement on the part of the researcher to decide how best to categorise the data. This information was entered (in summary form, along with verbatim quotes) alongside the outcome and contribution data, in the same row. Where a source provided evidence of only part of a CIMO (e.g. suggesting a particular mechanism was important without providing any insights into the contextual or intervention factors that sparked it), cells were simply left blank.

The analysis followed the analytical logic laid out in the diagram below.



Secondary data: Documents were reviewed by the country case study leads with the help of a research assistant. Programme leads compiled summary notes in Word. Evidence relating to *outcomes* was extracted into a second Microsoft Excel document review spreadsheet, as follows:

1. Documents were reviewed for insights on the outcomes generated by the programme.
2. This information was entered in summary form into the spreadsheet, coded according to which EQ the data related to.

Together, the primary and secondary data Excel sheets provided a catalogue of evidence enabling country case study leads to systematically and transparently assess the strength of evidence behind particular changes and identify how and why these changes were thought to have come about. Following discussions with the evaluation Steering Committee at Stage 2, the CIMO analysis has been embedded in the key findings sections in narrative form rather than presented as front-and-centre, in order to maximise the readability of the report.

3.4.4 Value for money analysis

A value for money (VfM) analysis was conducted as part of the Stage 3 country case studies, and integrated with the case study data collection and analysis. Given the summative stage of the evaluation, Stage 3 VfM analysis focused on cost-effectiveness – understanding the extent to which the investments made in the case study country had delivered value. At the country case study level, the focus was on assessing the cost-effectiveness of achieving priority outcomes. The cost-effectiveness of the overall investment made by BCURE partners in the case study countries will be made at a comparative level in the overview report.

Due to the nature of programme financial reporting it was not possible to identify the precise costs of programme activities. It was not a requirement for programmes to monitor the actual costs of activities and therefore the financial reports submitted to DFID did not provide an accurate picture. In many cases, the programmes had ended at the time the Stage 3 analysis was done, and so it was not possible for programme and financial staff to spend the time necessary to generate accurate data. As a result, the cost data was in most cases a rough estimation developed in consultation with BCURE programme staff.

Given the data limitations, it was not possible to conduct a robust quantitative VfM assessment, so qualitative judgements were made, were made through considering the following questions:

- Did the outcomes that were achieved justify the costs? Was the balance of investments across the priority outcomes appropriate?
- How institutionalised and/or sustainable were the reforms and outcomes observed?

3.5 Component 2. Literature review

A realist literature review (Punton et al., 2016a) was conducted during the early stages of the evaluation, in 2014–15.[7] The aim of the review was to provide a practical summary of recent evidence on what works to promote EIPM, in order to both contribute to the wider evidence base and begin developing CIMO configurations. The findings informed the emerging theory and the development of the first iteration of CIMOs tested in Stage 1. A light touch literature review refresh was conducted in 2017 in order to generate additional insights on the Stage 2 programme theory, and the insights incorporated into the final comparative report.

3.6 Component 3. Impact case study

A non-BCURE impact case study was conducted, in order to generate evidence on how capacity building for EIPM can lead to improvements in the quality of policy processes, the hoped-for ultimate impact of the BCURE programmes. This was designed to complement the BCURE programme evaluations through examining a non-BCURE capacity-building intervention that had been operating for a longer period of time,

[7] Punton et al., (2016a). Available from http://www.itad.com/knowledge-products/bcure-literature-review/

thus offering the potential to investigate how capacity building could contribute to changes in policy quality in the longer term.

The impact case study was the focus of an evaluability assessment and scoping process during the inception phase, detailed in the inception report. South Africa was selected as the country that most closely met the criteria. The study focuses on the Department for Planning, Monitoring and Evaluation (DPME), exploring the National Evaluation System (NES) as an example of a capacity support initiative that intervenes at organisational level to enhance evidence use in policymaking and has been established for some time (since 2011), providing an opportunity to investigate how capacity building can promote change in the longer term. The core research question for the impact case was: **How has DPME's support to the NES influenced evidence use and contributed to changes in the quality of policy processes?**

To answer this, the case study looked specifically at two experiences with line ministries. The first is the updating of the government of South Africa's early childhood development policy following a DPME-facilitated diagnostic review in which the Department of Basic Education had a leading role. The second experience is the evaluation of the Department of Trade and Industry's Business Process Services programme and changes in the programme design arising from the evaluation.

There were three main analytical strands to the impact case study: developing and testing CIMOs at the level of organisational change; researching the policy development process in order to provide insights into the concept of 'policy quality'; and exploring the interrelationships and dynamics between CIMOs and how they influence policy processes. The case study sought to explain how and why evidence produced through the evaluation/review of these policies/programmes was used in decision making. It also examined connections between evidence use and enhancement of policy processes in the two departments concerned.

The case study followed the same data collection and analysis methods as the Stage 2 programme evaluations, and took place during Stage 1 and 2 of the evaluation. It involved a review of relevant documentation as well as interviews in South Africa with DPME staff members, intervention participants, high-level stakeholders, civil society or other external stakeholders and service providers. In total 39 interviews were conducted in Stage 1 and 2, involving 32 unique interviewees. Data was analysed in the same way as programme evaluation interview data, as detailed in the Stage 2 synthesis report and annexes. The final evaluation also drew on insights from an evaluation of the NES, due to be finalised in 2018. This was not yet published at the time of writing the final evaluation, but insights were drawn from presentations provided at the SAMEA conference in 2017.

Table 5: Number of stakeholders consulted in impact case study

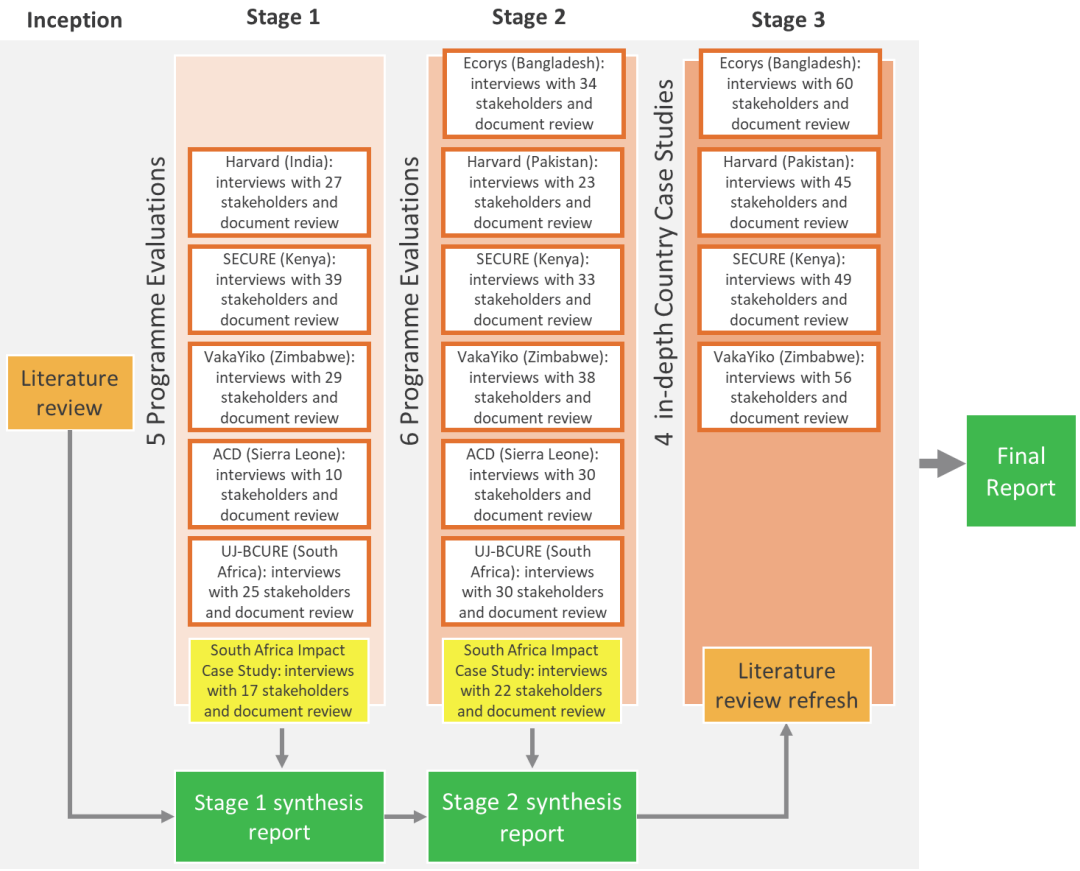
Category of respondent	Total stakeholders consulted for impact case study
DPME staff	8
Intervention participants	11
High-level stakeholders, e.g. senior leaders in national government; national research community; others	8
Civil society/other external stakeholders	5
Total	**32**

3.7 Component 4. Overall synthesis

The overview report brings together the findings from the full three years of evaluation outputs: the Stage 1 and 2 programme evaluations and Stage 3 country case studies, the literature review and impact case study, and the Stage 1 and 2 synthesis reports. It aims to draw generalisable conclusions about how and why different BCURE interventions have contributed to different patterns of outcomes in different contexts, producing an evidence-based set of refined CIMOs and a refined programme theory.

Figure 3 presents a summary of the data from various evaluation components, illustrating how this fed into this final report. In total, 528 stakeholders were consulted across six countries and over three years.

Figure 3. Summary of data feeding into the final evaluation report



The synthesis process involved:

1. **Initial calls and workshops** with the country case study leads, to identify common concepts, themes or metaphors that applied across the cases, and interrogate differences. This enabled patterns to be identified and helped reveal nuances in the findings.

2. **Using a synthesis database** to combine relevant evidence from across the four Stage 3 case study reports about the outcomes achieved and not achieved (EQ 1), BCURE's contribution to these outcomes (EQ 2) and how and why particular outcomes were and were not achieved (EQ 3).

3. **Conducting a realist synthesis** across the cases, exploring how and why different BCURE interventions contributed to different patterns of outcomes in different contexts (EQ 3), in order to produce evidence-based set of refined CIMOs. As well as the Stage 3 case studies, this process also drew on the Stage 1 and 2 synthesis reports, the impact case study report and the literature review / literature refresh. This process applied realist synthesis techniques and additional insights from meta-ethnography in order to draw out meaning in a systematic way (see Box 2). This was a highly analytical and creative process. It was undertaken by two members of the core team, which enabled cross-checking of coding and analytical decisions, and constant communication via Skype and email to help clarify, refine and challenge the analysis.

4. **Checking and validating emerging conclusions**, through reviewing case study reports and where necessary interview data, to ensure that the evidence used to support, refute or refine the hypotheses

underlying the findings was relevant and sufficiently rigorous to support the inferences made (the 'translation' step in meta-ethnography, see Box 2). The two lead researchers cross-checked each other's analysis and conclusions, and shared drafts with other members of the core evaluation team to further validate and nuance findings.

During step 2, a few broad and cross-cutting patterns emerged that appeared to explain incidences of success across the BCURE portfolio. These patterns were discussed within the evaluation team, and then systematically analysed by developing matrices that drew together relevant insights from across the sources, and applying the synthesis techniques described in Box 2. This analysis suggested the importance of three broad 'ways of working' when seeking to build capacity for EIPM, described in Section 5.

Box 2: Qualitative synthesis techniques

Realist synthesis is an iterative process of theory building. It aims to generate the best possible explanation of evidence through retroductive analysis: moving between inductively building theories, and deductively testing them, while (in line with the realist philosophy) acknowledging that evidence and the resulting theories will always be partial and incomplete. Retroductive analysis applies a range of techniques to draw out insights from data, including: (Michaelis and Westhorp, 2016; Pawson, 2006)

- **Juxtaposing** insights from one case study to make sense of an outcome pattern noted in another.
- **Reconciling** contradictory insights through unearthing contextual or implementation differences and showing how these lead to opposing outcomes.
- **Adjudicating** between contradictory findings from different cases, to unearth strengths and weaknesses in the original conclusions that may explain these contradictions.
- **Consolidating** different results into multi-faceted explanations of success.
- **Situating** different results in their contexts – e.g. by exploring how one mechanism might operate in context A while another may operate in context B.

Meta-ethnography has much in common with realist synthesis. It is also an interpretive synthesis method, involving the transfer and translation of ideas, concepts and meanings across different sources (Noblit & Hare, 1988). Two of its steps were helpful as additional techniques for the synthesis:

- **Determining how evidence was related**: identifying points of comparison or opposition within the case studies, and identifying 'lines of argument' – inferences that cut across cases – through "*comparing and sorting interpretations, examining similarities and differences, and then integrating or framing these within a new interpretation"* that applied across cases (Pope et al., 2007).
- **Translation**: periodically revisiting case study reports and interview data to attempt to 'translate' evolving concepts or theories back into the source data, checking to see how far they accurately reflected case study findings, and scrutinising conceptual differences.

Throughout the overview report, insights on 'what worked, for whom, and why' have been drawn out. These represent 'empirical' CIMOs, which explain specific **outcomes** (O) from across the BCURE projects in terms of the **mechanisms** (M) that were (or were not) sparked by resources provided by BCURE, and the **context** (C) and **intervention** (I) factors that enabled or constrained the mechanisms. In the conclusions, these empirical formulations are brought up to the level of middle-range theory,[8] representing our final tested theory about what works to build capacity for EIPM, for whom, and in what circumstances.

3.8 Judging strength of evidence and extent of contribution

'Strength of evidence' relates to the internal validity of the evaluation findings. Our aim through the Stage 3 evaluation was to achieve a *sufficient degree of confidence* about the extent to which priority outcomes have occurred (EQ 1), BCURE's level of contribution to the outcomes (EQ 2) and our theories (CIMOs) about how

[8] This is theory that is *"detailed enough and 'close enough to the data' that testable hypotheses can be derived from it, but abstracted enough to apply to other situations as well"* (Wong et al., 2013).

and why BCURE contributed or failed to contribute (EQ 3).[9] Confidence in our conclusions about outcomes, contribution and CIMOs is underpinned by three broad considerations:

1. The extent of triangulation across stakeholders, participants/non-participants, and/or data sources. Within BCURE, triangulation has been pursued on several levels:
    - **Within interviews, by asking for examples.** If a stakeholder claims to have observed an outcome, confidence that this is true is increased if they are able to give specific examples.
    - **Across stakeholders and types of stakeholders**. Confidence that an outcome has occurred is stronger if more people, across different groups, claim to have observed it. Where possible, this has included seeking out and comparing insights from programme participants with non-participants, who have less of a stake in the programme being perceived as successful, and who, due to their position, have independent insights that provide corroboration and contextual information.
    - **Across data sources**: We have sought to triangulate insights from primary data collected through interviews with M&E data collected by the programme, and where possible with documents (e.g. policy documentation) produced by BCURE participants.
2. **A consideration of the position, knowledge, analytical capacity, reflexivity, and potential biases of primary informants.** In line with our realist approach, sampling decisions were purposively and iteratively guided by our (existing and emerging) *theory* about the outcomes we expected to observe, and how and why these outcomes were expected to come about. Stakeholders are therefore not considered in terms of homogenous categories (participants / non-participants), but as individuals positioned in unique ways in relation to the programme, with different levels of knowledge, capacity and reflexivity, and different incentives that may lead to bias. Weighing the strength of evidence requires a consideration of these issues, rather than simply considering the number of respondents who confirmed a particular outcome or CIMO. For example:
    - Different people can be expected to know different things about an expected outcome or change process. In some cases, only a small number of people are likely to know about an outcome, BCURE's contribution, and how / why it happened. Weighing the strength of evidence requires the evaluators to judge whether those who can be expected to know about the issue have confirmed that things happened in a certain way.
    - Different respondents have different levels of capacity (and interest) in scrutinising how and why something happened – particularly when this requires them to consider why they themselves have (or have not) changed their attitudes or behaviours – and this affects the weight that should be given to their responses.
    - Different stakeholders will have different incentives which may lead to biased responses; most obviously an incentive to 'tell the evaluator what they want to hear' in order to paint the project in a positive light and potentially secure future funding, leading to confirmation bias**.**
    - The position of a respondent in relation to the programme gives them a particular perspective which needs to be considered, overlapping with all of the above considerations. An external sectoral stakeholder may be able to provide important independent insights about broader political economy issues, but may not know much about the specific individuals or teams who took part in the programme (and therefore their opinions should be weighed accordingly). A senior civil servant may have good insights into outcomes but may be unwilling to speak openly about the realities of incentives and power structures within their ministry, and although they may not have participated directly in the programme they still stake in its success which implies the need to mitigate possible bias.

[9] This draws on thinking from process tracing and contribution analysis. Process tracing in particular offers useful insights into how to qualitatively weigh evidence in order to 'increase our confidence' that an intervention had an impact in a particular way. See: Befani, B. and Mayne, J. (2014). Process Tracing and Contribution Analysis: A Combined Approach to Generative Causal Inference for Impact Evaluation. *IDS Bulletin*, *45*(6), 17–36. http://doi.org/10.1111/1759-5436.12110

The evaluators considered these issues both during the sampling process (when making decisions about who to interview), and during the interview write up and analysis (taking note of issues on the analysis spreadsheets in order to feed these considerations into the write up).

3. **A consideration of the broader context.** At Stage 3, the evaluation took a more explicit look at the broader political economy factors that enable and constrain EIPM in the countries and sectors under examination, and which provide opportunities and risks to the programme. This has provided more detailed insights into the contextual dynamics of BCURE country programmes, helping ensure that explanations of change are grounded in an understanding of the political context and are not over-reliant on the explanations of programme participants. This also helped identify other (non-BCURE) explanations of change, in order to help guard against over-attributing change to BCURE.

These three considerations were used to develop a qualitative approach to assessing the strength of evidence, described in the table below. This is not a rigid framework, but a way to ensure the evaluative judgements were made systematically and are comparable across the four case study reports.

Strength of evidence	EQ 1	EQ 2	EQ 3
Strong evidence	High level of confidence that the outcome occurred…	High level of confidence that BCURE contributed to the outcome…	High level of confidence that the outcome occurred / did not occur as a result of *x* mechanism, operating in *y* context and as a result of *z* features of the intervention…
	• …Based on a good degree of triangulation a) within interviews, b) across stakeholders and types of stakeholders, and/or c) across data sources • …Taking into account the position, knowledge, analytical capacity, reflexivity, and potential biases of primary informants • …and also taking into account what we know about the broader context through the PEA insights		
Some evidence	More confident than not that the outcome occurred…	More confident than not that BCURE contributed to the outcome…	More confident than not that the outcome occurred / did not occur as a result of *x* mechanism, operating in *y* context and as a result of *z* features of the intervention…
	…But confidence is reduced by: • Shortcomings with regards to triangulation, and/or • Concerns that the position, knowledge, analytical capacity, reflexivity, and potential biases of primary informants lowers the reliability of evidence, and/or • What we know about what is happening within the broader context		
Limited evidence	Low level of confidence that the outcome occurred, given that…	Low level of confidence that BCURE contributed to the outcome, given that…	Low level of confidence that the outcome occurred / did not occur as a result of *x* mechanism, operating in *y* context and as a result of *z* features of the intervention, given that…
	• …Evidence comes from a small number of sources with limited triangulation, and/or • …there are major concerns that the position, knowledge, analytical capacity, reflexivity, and potential biases of primary informants lowers the reliability of evidence, and / or • … there are contradictory insights into what is happening within the broader context		

Judging extent of contribution
In relation to EQ 2, a judgement was made regarding the significance of the programme's contribution to change. This represents a qualitative judgement on the part of the lead evaluator, based on a consideration of evidence collected relating to other factors that may have contributed to change.

Contribution rating	Details
Crucial contribution	Evidence that programme made a crucial contribution to observed change; i.e. change would not have happened without the programme. OR observed change is directly attributable to the programme
Important contribution	Evidence that programme made an important contribution alongside other factors
Some contribution	Evidence that programme made some contribution alongside other factors, but was not the most important cause
No contribution	Evidence of no contribution, or no improvement evident
Insufficient evidence	Insufficient evidence to make an assessment

3.9 Stakeholder engagement throughout the evaluation

The BCURE evaluation has been designed and implemented in close collaboration with the DFID evaluation Steering Committee, through regular meetings and calls, as well as numerous internal approach papers which offered an opportunity for DFID to review and comment on emerging design choices and suggested report structures. This regular engagement has facilitated annual revisions to the design in order to ensure the evaluation is meeting DFID's needs, particularly at Stage 3 where a substantial redesign was conducted (described above). The Steering Committee was also consulted on the selection of priority outcomes and CIMOs to test at Stage 3, based on the issues and questions most relevant to the design of future programmes.

BCURE partners have also been engaged at various points throughout the evaluation. Annual BCURE learning events offered an opportunity for the evaluation team to share emerging findings and interim analysis, with comments from partners fed into synthesis reports. In-country workshops with project partners provided an opportunity to hear the views of implementation teams and test CIMOs against their understanding of how and why change was (or was not) happening. At Stage 3, draft country ToCs were also shared and discussed with BCURE partners, and revised accordingly. Where possible during country visits, the evaluation leads also conducted debrief interviews or workshops with project staff, to share emerging findings at the end of the fieldwork, answer partner questions, and sense-check interpretations. Finally, draft programme evaluation and country case study reports were shared with partners to provide an opportunity for comments before the reports were finalised. These reports are internal, in order to protect the confidentiality of respondents and the relationships of BCURE partners with government stakeholders. However, synthesis reports and other publicly available evaluation products have been shared with interviewed stakeholders.

3.10 Ethics

The key ethical issue faced in the evaluation was protecting and managing the confidentiality of government documentation and stakeholder views at the local level. A number of the BCURE partners were operating at a high level within government and as such had access to government policy processes as they unfold. Access to these processes and the actors involved was been navigated with the close collaboration of the BCURE partners, in order to avoid the evaluation negatively impacting the relationships that BCURE providers have worked hard to develop.

The evaluation team sought to collect data in an appropriate and respectful manner, taking into account cultural and ethical concerns. Where possible, potential interview respondents were contacted several weeks in advance to enable the evaluation to fit into busy government schedules. Access to government institutions was facilitated by BCURE partners and national consultants, who had local knowledge about the protocols and etiquette involved in accessing stakeholders at varying levels of seniority, and who briefed international consultants on this in advance. Field trips were scheduled to allow sufficient time for researchers to be able to change their plans in order to fit in with the fast-changing schedules and commitments of government stakeholders. Researchers were also respectful of participants' time, and frequently cut interviews short or changed venues to enable stakeholders to participate. Researchers were also mindful of questions that might be inappropriate in particular contexts.

We ensured informed consent was obtained from individuals before commencing data activities, with consent obtained at the beginning of interviews to record the discussion and to use the insights gained in our reports (see Annex 7). Unique anonymous interview codes were attached to each transcript and referenced in the country case studies. Where the content of quotes had the potential to identify an individual, this information was removed.

The BCURE country case studies are confidential reports viewed only by DFID and by the programme teams. The overview report aims to reflect on findings at a higher level of abstraction, allowing us to avoid detailed reporting on sensitive issues relating to particular government processes.

3.11 Evaluation team

The evaluation was undertaken by a core team from Itad, in partnership with Stellenbosch University.

Core team members

Isabel Vogel (Itad Associate)	Team Leader and lead evaluator for Kenya case study
Rob Lloyd (Itad Associate Director)	Project Director
Melanie Punton (Itad Senior Consultant)	Lead evaluator for Zimbabwe case study, support to Bangladesh case study, methods advisor
Gregory Gleed (Itad Consultant)	Lead evaluator for Pakistan case study
Joe Bolger (independent consultant)	Lead evaluator for Bangladesh case study, support to Pakistan case study, lead researcher on impact case study
Teresa Hanley (independent consultant)	Lead evaluator for Sierra Leone case study (Stage 1 and 2), support to Kenya case study
Babette Rabie (Stellenbosch University)	Lead evaluator for South Africa case study (Stage 1 and 2), support to Zimbabwe and impact case studies
Fanie Cloete (Stellenbosch University)	Lead evaluator for India and Zimbabwe case studies (Stage 1), advisory support

Country case studies were supported by national consultants Maheen Sultan (Bangladesh), Alfred Odour and Susan Mathai (Kenya), Rafiq Jaffer (Pakistan), Munhamo Chisvo (Zimbabwe), Andrew Lavalei (Sierra Leone), Benita Williams (South Africa) and Milindo Chakrabarti (India). Research assistance was provided by Alexina Jackson, Greg Smith, Verdiana Biagioni, Louise Horton and Katharine Hagerman. Communications support was provided by Clare Gorman and Emmeline

3.12 Limitations to the synthesis

The evaluation team was able to work freely and without interference, and there are no conflicts of interest to report.

The Stage 3 evaluation attempted to address a number of limitations identified at Stage 2. While this was successful to a large extent, certain issues proved difficult to address, giving rise to important lessons for future realist evaluations and evaluations of EIPM capacity development programmes.

The realist evaluation approach has been challenging to implement across the complex BCURE programme. In particular, we have faced challenges with 'breadth vs depth', including how to systematically prioritise outcomes and theories to assess within the limited time available for interviews. The Stage 1 and 2 evaluations generated a large number of theories (CIMOs) about how and why BCURE might be contributing to change at individual, interpersonal, organisational, institutional and policy levels. At Stage 2, it became clear that it was not possible to systematically test theories across the whole BCURE theory of change with the resources available for the evaluation. This was mitigated at Stage 3 through conducting a smaller number of more in-depth case studies, and prioritising a smaller number of outcomes and CIMOs for investigation. Developing country-level ToCs rather than relying on an overarching ToC helped identify case-specific outcomes and CIMOs that were less well-evidenced through earlier stages of the evaluation, and which were most important for achieving longer-term outcomes. This approach proved largely successful, and highlights the importance of realist evaluations prioritising the most interesting and important causal links in enough depth to draw useful insights, rather than trying to investigate everything. Case-specific theories, rather than (or as well as) a single overarching theory, can help facilitate this, through building an in-depth understanding of how and why a programme is expected to unfold in a specific case.

It has also been challenging to encompass complexity within the CIMO framework, including features of the macro political context and how they give rise to or inhibit mechanisms of change. There is a risk that CIMOs become overly linear explanations of how and why change happens ('this intervention feature, in this context, sparks that mechanism to lead to this outcome'). This was mitigated by presenting the final CIMOs in a more narrative way, which allowed the nuances and interconnections to be unpacked. The Stage 3 CIMOs also contain multiple features of context and multiple mechanisms, illustrating how these work together to lead to outcomes.[10]

In Stages 1 and 2 of the evaluation, it proved much easier to identify 'micro' features of context (e.g. around the characteristics of trainees) than 'macro' features (e.g. around the nature of government systems, the influence of power, politics and high-level incentives). This was mitigated through incorporating a specific PEA step in the Stage 3 methodology, which significantly enhanced the richness of the analysis. However, while the interviews provided a wealth of insights into the risks and opportunities that the context posed for EIPM and the programme's desired outcomes, it was not possible to gain insights into certain important issues likely to affect evidence use, including actual distribution of power and decision making, and some of the individual and organisational incentives that affect evidence use by senior government stakeholders. This suggests the importance of building in an explicit PEA lens from the outset in future realist evaluations working in government contexts, and considering how the evaluation design and team can be structured to best gain access to information on power distribution and incentives.

The evaluation has by necessity relied on interview data for evidence of outcomes, and there is a real risk of positive (confirmation) bias of respondents. With some exceptions, BCURE project monitoring systems were not set up to capture evidence of outcome-level change (including behaviour change and changes in policy processes or content). This has proved a major challenge for the evaluation, suggesting the importance of ensuring future programmes build monitoring systems that go beyond measuring outputs such as self-reported increases in knowledge and skills. The evaluation explored the possibility of conducting

[10] This follows the example of Leavy, Boydell and Mcdowell (2017).

large-scale surveys to capture insights from a broad cohort of participants, but given the high risk of low response rates this was not pursued. As a result, the evaluation has relied primarily on qualitative interviews with a select number of participants in order to provide evidence of longer-term outcomes. This carries a risk of confirmation bias, given the power dynamics of interviewing government stakeholders in low and middle-income countries. There is also the risk that participants may genuinely feel that the programme contributed to a positive outcome, when in fact other factors were more important – and this risk is heightened due to the complexity of the interventions, which makes it challenging to unpack contribution. We attempted to mitigate this at Stage 3 as follows:

1. **Triangulation:** We aimed for no more than 60% of the sample to consist of project participants and programme staff, with the remainder consisting of knowledgeable non-BCURE participants. Increasing the number of non-participant interviews helped to triangulate insights from project participants with the perspectives of individuals with less stake in the programme and potentially less incentive to tell the evaluators what they felt we wanted to hear.
2. **Conducting a more in-depth investigation into priority outcomes, and identifying and testing non-BCURE influences of change.** Focusing on a small number of priority outcomes enabled us to interrogate stakeholder testimony in more depth, helping us gain more detailed insights into what had happened and what had enabled or inhibited change. Our PEA exercise provided insights into country and sectoral contextual dynamics, helping ensure that explanations of change were grounded in an understanding of the political context, were not over-reliant on the explanations of programme participants, and were fair to programmes working in challenging settings. This helped to guard against over-attributing change to BCURE, as well as contextualising shortfalls in programme achievements.
3. **Dedicating more resources to finding monitoring and other documentary sources in order to triangulate interview data.** This included policy documents or government guidance that would help us validate stakeholder testimony about improvements in evidence access, appraisal and use. While this had some success and most case studies were able to view at least some documentation, there were ongoing challenges in accessing this data as the majority of stakeholders were unable to share internal government documents. This challenge was somewhat mitigated through interviewing a wide range of participants, and where possible their colleagues and managers, to triangulate insights.

It proved very challenging to secure interviews with government officials across all four settings – both participants and non-participants – particularly in Bangladesh and Pakistan. Challenges in securing and conducting interviews was a result of high workloads, adverse weather, security concerns (in Pakistan), and also the fact that most BCURE projects had largely finished activities in both settings, providing limited incentive for participants to volunteer their time. These challenges were mitigated through dedicating substantial efforts to contacting and following up with respondents, and through extending the length of country visits; however, in a number of cases the interviews were very short and it was only possible to explore a small number of outcomes and theories. This is reflected in the depth of analysis and strength of conclusions drawn in the country case study reports, and subsequently this overview report.

It has been challenging to ensure consistency of data collection and analysis across a diverse programme team. Time and budget constraints meant it was challenging to train the team comprehensively on the evolving programme theory, the principles of conducting realist interviews, and the approach to analysing data in a realist way. This created issues with ensuring a consistent approach to testing CIMOs and analysing interview data across the cases. At all three evaluation stages, we have attempted to mitigate this through a team workshop prior to data collection, involving a full introduction to the programme theory and basic training on realist interviewing and analysis. Programme leads provided training in-country to national evaluators prior to data collection, and additional analysis was conducted at synthesis stage by the team leader and methodological lead to capture insights that may have been missed during the case study analysis. At Stage 3, we also revised the team structure so that country visits were conducted by two core team members rather than one, which helped improve consistency across the cases. However, our major reflection is that realist evaluations require a different approach to team structuring and capacity building. Realist interviews and analysis require team members to have an in-depth understanding of not only the methodological approach, but the theory that the evaluation is trying to test. In order to ensure consistency

and understanding, a realist evaluation requires a more participatory model, which involves in-depth and ongoing engagement and capacity building.

Finally, the evaluation draws on evidence from only six of the 12 BCURE countries, and the short time frame of the programme limits the potential to record longer-term results. The evaluation is limited in what it can say about how BCURE worked across all of its settings, because its focus on six country case studies means it has not captured the full range of outcomes across the whole portfolio. Given the country-level focus, it also does not capture outcomes from the international and regional networking components that were part of various projects.
Finally, the programme was relatively short given its aim to generate systemic change in government settings – the shortest programme, in Bangladesh, had only two years of implementation time. This has limited the ability of the evaluation to identify longer-term results.

4 Programme theory and CIMO refinement

This section details the evolution of our programme theory and CIMOs from Stage 1 to Stage 3, documents the changes made and the rationale behind these changes, and presents our refined programme theory at the end of the evaluation. The Stage 2 Synthesis Annex contains further information about the evolution of theory from Stage 1 to Stage 2.[11]

Stage 1 BCURE ToC: Unpacking 'capacity development' to create a unifying framework

The six BCURE projects were highly diverse, taking different approaches to enhancing skills and systems for evidence use, in complex government contexts. As BCURE did not have a portfolio-level ToC, the evaluation developed an initial common theory of change (CToC) through synthesising the original project ToCs with key insights from the literature. The Stage 1 CToC unpacked capacity development into four levels of capacity change (See below),[12] which helped to bring the diverse BCURE approaches into a unifying framework for the evaluation. The four levels conveyed the concept of capacity development as multidimensional, and capacity as a function of different factors and processes working together and reinforcing each other at different levels. The BCURE ToC at Stage 1 stated that multidimensional change across these four domains would contribute to routine change in the use of evidence by government, in turn contributing to improved *quality of policy development processes,* as the overall impact.

Four levels of change

1. *Individual level:* individual behaviour (decisions and actions) in relation to EIPM, and the skills, knowledge, motivation, attitudes, commitment, values and personal incentives that affect this.
2. *Interpersonal/network level:* the relationships between individuals and groups that affect evidence interpretation and use, including formal and informal communities (or networks) of individuals or organisations.
3. *Organisational/government* level: the systems, policies and procedures, practices, culture or norms within a governmental organisation that exist above the level of individual actors, and which incentivise, support (or inhibit) evidence access, appraisal and application in decision making. This includes 'system-level' factors within government that affect EIPM, such as national or sub-national laws, policies, regulations,

[11] Available here: http://www.itad.com/reports/building-capacity-use-research-evaluation-bcure-realist-evaluation-stage-2-synthesis-report/

[12] There are many definitions used in the literature to describe levels of capacity change. We adapted DFID's definitions from the 2010 'How to Note on Capacity Building in Research' (DFID, 2010). This document uses 'institutional' to denote 'changes in the rules of the game'. Other readers may interpret 'institutional' to mean 'systemic' or 'environmental' change. We opted to consider the government system as falling within a broadly conceived 'organisational change' category because organisations within the government system are bound by common, cross-cutting rules, incentives and procedures. This means that 'institutional' change then encompasses all non-governmental influences within the wider environment. However, we recognise that the boundaries between the levels of change are fuzzy and dynamic.

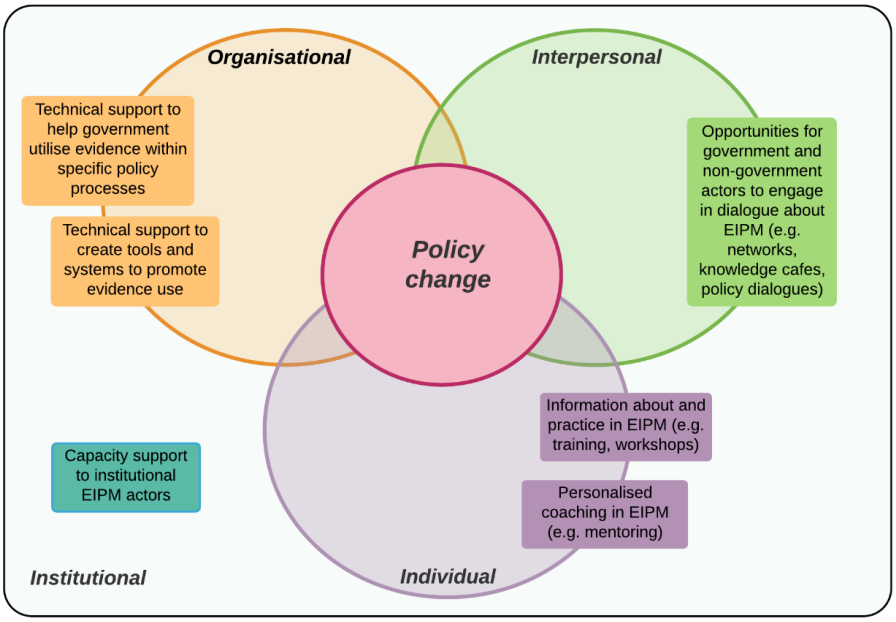
governance systems and 'institutional rules of the game'. Our definition of 'government' includes government administration and parliamentary scrutiny functions (including elected opposition politicians).

4. *Institutional level:* the broader enabling environment for evidence use *outside* of government, including the role of external actors, such as international donors, civil society and the media, and the influence of external factors such as crises, global events and socioeconomic change, as well as broader societal factors that influence EIPM, such as culture, norms, collective beliefs, attitudes and values. This includes the institutional role of the BCURE partners themselves within their national contexts.

Stage 2 ToC: Unpacking 'evidence use' and EIPM as a system working on multiple levels

The Stage 2 evaluation confirmed our theory that changes emerging at different levels (e.g. individual skills and organisational systems) seemed to reinforce each other, and that changes at different levels were required to make progress towards improvements in the quality of policy products and processes. The BCURE interventions worked through different 'entry points' at different levels. Some projects initially targeted *individuals* with information and opportunities to practise skills, others provided spaces for dialogue between different *groups* of stakeholders, others delivered technical support to *organisational* systems and processes and/or develop the capacity of *institutional* actors to promote EIPM. This led us to formulate the programme theory emerging from Stage 2 findings as a set of interlocking domains, with entry points at individual, interpersonal, organisational and institutional levels.

Figure 3: Different entry points of the BCURE interventions



The Stage 2 programme theory is described below, and represented in diagrammatic form in Figure 6.

Stage 2 programme theory and CIMOs

When the programme 'entry point' is through interventions targeting *individuals*…

- Providing information about EIPM (its importance, and how to access, appraise and apply evidence in decision making), alongside opportunities to practise skills, generate self-efficacy (a feeling of 'now I know how') and lead to behaviour change when training is directly relevant, there is management support and training comes at the 'right time' for the organisation (CIMO 1)
- Coaching provides *encouragement*, which generates or embeds a feeling of self-efficacy ('now I know how'); *contacts and sponsorship* that give access to useful networks; and *advice and a guiding hand* that promote understanding and builds confidence. This can result in participants changing their behaviour in relation to EIPM where they have either personal motivation or organisational incentives to do so. Success depends on coaching being driven by clear objectives based on participants' needs, and the coach having the right interpersonal and professional qualities to provide for these needs (CIMO 2).
- Facilitated spaces for dialogue and collaboration can enable advice and sharing of perspectives to generate knowledge and influence attitudes about EIPM, including learning about what others have done when facing similar challenges. This is made possible where interventions bring together diverse groups of people with relevant interests, and provide space to share challenges, in a context of a positive wider discourse in support of EIPM. However, this learning may be put into only use if there are

existing direct opportunities to do so, although spaces for dialogue potentially create a conducive context for other interventions to stimulate behaviour change at a later stage (CIMO 3).

- Providing individual-level support (such as training or coaching) in a sensitive, collaborative way can provide a 'foot in the door' for BCURE partners, generating permission and buy-in for them to begin implementing organisational reforms – this could be a particularly important 'way in' in contexts where it is not possible to start working directly at organisational level, for example where access to government is difficult to secure (CIMO 5).

When individuals began using evidence more in their day-to-day work, this can catalyse organisational change as follows:

- When a sufficient number of individuals (including some with leadership roles) begin accessing, appraising and applying evidence more in their work, this can 'filter up' and lead to higher-level recognition of the value of an evidence-informed approach – through senior staff seeing and being impressed by good-quality evidence products and through these products feeding into senior decision making processes and improving them (CIMO 6).
- When individual support influences individuals in mid-level roles, who are committed and passionate and who have supportive senior management, they can formally cascade their learning through introducing new ways or working and new structures and processes within their organisations (CIMO 7).

When the 'entry point' is through interventions attempting to improve *interpersonal* links or relationships…

- Facilitated spaces for dialogue (e.g. between policymakers, researchers, civil society and citizens) can create and strengthen connections or generate a sense of closeness and trust, resulting in new and improved relationships. This is more likely where open, informal dialogue is enabled, where the 'right' composition of people are in the room, and in contexts where existing networks are weak or dysfunctional but there is a positive wider discourse in support of EIPM. Where participants have the motivation or opportunity to utilise new relationships, they can be used to share information or advice, or can lead to new organisational collaborations (CIMO 4).

When the 'entry point' is through interventions at *organisational* level…

- Providing technical support to co-produce tools or systems that facilitate staff to use evidence more effectively, where this is done in a collaborative and innovative way, can generate good examples that 'showcase' the value of evidence for quality, performance and delivery. These 'showcases' provide user-friendly decision support tools that help individuals use evidence, but also build understanding and buy-in among senior staff about the value of evidence for decision making, resulting in examples 'diffusing' out to inspire new reforms elsewhere (CIMO 8).
- Where there is pressure to improve performance from senior levels and where an external partner has established trust through previous activities, this can enable an 'accompaniment' mechanism: high-level stakeholders give partners the permission to provide ongoing, tailored support to help them embed EIPM. This can lead to uptake of recommendations from processes facilitated by the partner, adoption of tools or systems, and possibly the emergence of an internal unit to 'own' and 'champion' EIPM (CIMO 9).
- Providing technical support to co-produce tools or systems that facilitate staff to use evidence more effectively can spark a high-level decision to formally adopt the tools or systems to help standardise EIPM within the organisation. This is more likely when they link to other government procedures and are backed by sufficient authority. Adoption can be on a small scale (e.g. adopting templates), but, in a context where there are high-level government 'owners' of EIPM, adoption can also be large scale (e.g. adopting a comprehensive policy and planning system to promote, embed and monitor the quality of evidence use throughout the policy cycle and into the future) (CIMO 10).

Organisational-level change can then filter down to influence individual behaviour through:

- Tools or systems to promote EIPM sparking a *facilitation* mechanism – providing practical assistance enabling people to do their jobs better / more easily. This results in the system or tool being used, and (potentially) increasing the value of evidence through demonstrating the benefits it can bring (CIMO 11).
- Tools or systems that involve positive or negative incentives to adopt EIPM behaviours sparking a *reinforcement* mechanism, in which positive incentives or risk of negative consequences influence behaviour, and lead to individuals deciding to change the way they access, appraise or apply evidence in decision making (CIMO 12).

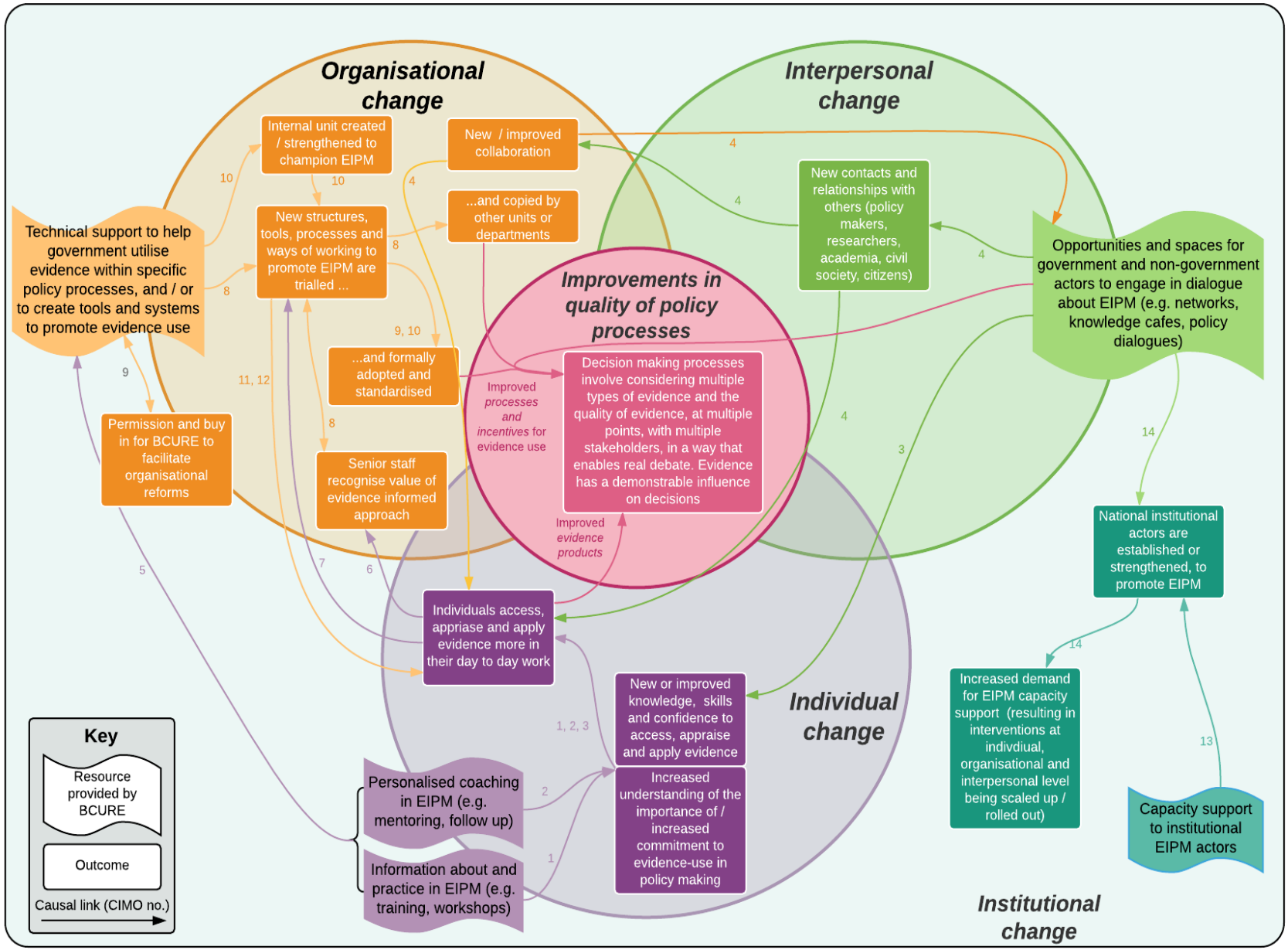
When the 'entry point' is through interventions at *institutional* level…

- Supporting local organisations to deliver EIPM capacity-building activities (directly through organisational capacity support, and/or indirectly through providing opportunities for national partners to 'learn on-the-job'), can strengthen organisational capabilities through 'learning by doing.' This can result in the establishment or strengthening of national institutional actors, which can act as a 'hub' for EIPM, are capable of running successful programmes to promote it and are potentially able to continue supporting it once the programme has ended (CIMO 13).
- Where local organisations successfully deliver programme activities and/or explicitly aim to build relationships with government departments and other EIPM actors, this enables partners to 'relate and attract' – providing exposure to new collaborators. This can lead to increased demand for partners to provide capacity building support for EIPM from new actors not originally targeted by the programme – which can provide a crucial entry point where there are sensitivities around influencing government decisions, and hence where it is difficult for 'outsiders' to gain entry to government organisations (CIMO 14).

Capacity change at individual, interpersonal, organisational and institutional level combines to contribute to improvements in quality of policy processes through:

- Improving *evidence products* (i.e. how evidence is prioritised, analysed, visualised and presented in briefing notes, evaluations etc.), which feed better quality or additional types of evidence into decision-making processes.
- Improving *processes and incentives* for evidence use – facilitating and incentivising decision makers to participate in policy development processes that involve explicit consideration of evidence.

Figure 4. Diagram representing BCURE programme theory at the end of Stage 2



Stage 3 design phase: Unpacking 'evidence use' and EIPM as a system, and defining 'policy quality'

At the end of Stage 2, it became clear that the evaluation needed to gain a deeper understanding of key political economy dynamics in order to explain BCURE's emerging outcomes. In order to engage with this, at Stage 3 we developed country-specific ToCs, drawing on the Stage 2 CToC and insights from the first two years of the evaluation, to provide a more concrete and contextualised framework for the country case studies. This allowed us to explore the critical political economy dynamics affecting the observed and hypothesised causal links. The synthesis level CIMOs were mapped on to these country-specific ToCs, providing a common set of theories to explore in different contexts.

The definitions of 'policy quality' and overall impact were also revised, to enable an assessment of progress towards impact in the final evaluation. We unpacked our working definition of 'policy quality' from Stage 2 into four priority outcomes and an impact statement, to reflect *embedded, transparent, conceptual and instrumental* uses of evidence. Our definition proposed that, for evidence use to promote critical thinking, a decision-making process needs to be *transparent* about the limitations of evidence by engaging explicitly with diverse perspectives and values and multiple types of evidence, and it should be transparent about the extent of evidence and its quality. In this way, productive debate and discussion on the issues raised by evidence can be encouraged and evidence is likely to have a demonstrable influence on the decisions made, whether conceptual or instrumental. However, occasional uses of evidence are not enough to achieve the impact. A key part of BCURE's intended impact was for evidence use to become *embedded* in decision-making routines, supported by organisational systems and incentives to use evidence. These concepts linked to emerging DFID thinking on measuring the VfM of evidence-into-use interventions, and also linked to key insights in the EIPM literature (see Box 3). In this framing, the evaluation recognised that dimensions a-d may also contribute to *strategic, tactical* and *political* uses of evidence, as well as potentially evidence *misuse*, which would fall short of BCURE's anticipated impact.

Box 3: Insights from the literature: Understanding 'evidence use' in policy processes

The BCURE literature review highlighted the different ways that evidence is used in policy design, decision making and implementation. Weiss (1972, 1980, 1982) emphasised that **instrumental** use of evidence, where specific evidence directly shapes policy choices, is only one way – and is often quite rare. More common is where evidence generates a slow 'enlightenment' as concepts and theories from research gradually percolate through society, *'coming to shape the way in which people think about social issues'*. This was labelled **conceptual** use of evidence by Nutley et al. (2007). However, evidence may just as frequently be used to justify or refine a position that has already been reached, which can be thought of as **strategic**, **tactical** or **political** use. There are also examples of unambiguous **misuse**, when poor quality findings are used, or tactical use of evidence intentionally justifies a bad practice (Nutley et al., 2007). Finally, there are examples of **over-use**, where a set of findings may become a new fad and be applied uncritically and wholesale.

Several sources in the literature review emphasise that evidence itself is not a neutral product – first because it reflects pre-existing views, values and beliefs of researchers and commissioners involved in producing it; and second because it rarely points to an obviously optimal solution, so that contestation over its meaning is inevitable (see for example, du Toit, 2012; Waldman, 2014). This suggests the importance of considering the nature of the decision-making process, and how different evidence sources and stakeholder perspectives are consulted and interpreted.

Table 1. Dimensions of policy quality

Dimension a	Government officials routinely consider a range of evidence and the quality of evidence when developing policy products (embedded use)
Dimension b	Appropriate policy development processes engage with evidence from diverse stakeholders and multiple perspectives (transparent use)

Dimension c	Routine evidence use is *facilitated by evidence tools*, which allow officials to access, identify and critically appraise the evidence base and apply it to decisions, being transparent about the evidence base behind decisions (transparent use)
Dimension d	Routine evidence use is reinforced, incentivised and monitored through processes and standards, supported by senior managers (embedded use)
Impact	Together, a–d are expected to contribute to conceptual and in some cases instrumental use of evidence: evidence indirectly shapes the way in which people think about social issues and in some cases has a demonstrable influence on the decisions made … and ultimately evidence-informed decisions are implemented

During the Stage 3 design phase, the decision was made in consultation with the Steering Committee to prioritise a focus on the 'longer-term' theories crucial to explaining how and why the projects did and did not contribute towards the impact. This also reflected the need to limit the investigation to a smaller number of CIMOs in response to the challenges of breadth vs depth noted in Annex 3.12. Individual-level change was only investigated insofar as it helped to contribute to longer-term, more routine shifts in evidence, or contributed to changes at the organisational level. Theories relating to interpersonal change were not prioritised in part because the Steering Committee was less interested in understanding change at this level, and in part because Stage 2 suggested change at this level was important as part of the context that enabled individual and organisational change, rather than a standalone outcome. Finally, most changes in the institutional domain were beyond the scope of the evaluation to investigate in depth, and also relatively minor part of most project activities, so theories relating to this were only investigated in some contexts.

The table below details which CIMOs were prioritised for testing at Stage 3.

Stage 2 CIMOs	Status at Stage 3
CIMO 1. Self-efficacy Providing information about EIPM (its importance, and how to access, appraise and apply evidence in decision making), alongside opportunities to practise skills, generate self-efficacy (a feeling of 'now I know how') and lead to behaviour change when training is directly relevant, there is management support and training comes at the 'right time' for the organisation	Not prioritised for investigation
CIMO 2. Coaching Coaching provides encouragement, which generates or embeds a feeling of self-efficacy ('now I know how'); contacts and sponsorship that give access to useful networks; and advice and a guiding hand that promote understanding and builds confidence. This can result in participants changing their behaviour in relation to EIPM where they have either personal motivation or organisational incentives to do so. Success depends on coaching being driven by clear objectives based on participants' needs, and the coach having the right interpersonal and professional qualities to provide for these needs	Not prioritised for investigation
CIMO 3. Learning from similar challenges Facilitated spaces for dialogue and collaboration can enable advice and sharing of perspectives to generate knowledge and influence attitudes about EIPM, including learning about what others have done when facing similar challenges. This is made possible where interventions bring together diverse groups of people with relevant interests, and provide space to share challenges, in a context of a positive wider discourse in support of EIPM. However, this learning may be put into only use if there are existing direct opportunities to do so, although spaces for dialogue potentially create a conducive context for other interventions to stimulate behaviour change at a later stage	Not prioritised for investigation
CIMO 4. Facilitated spaces for dialogue Facilitated spaces for dialogue (e.g. between policymakers, researchers, civil society and citizens) can create and strengthen connections or generate a sense of closeness and trust, resulting in new and improved relationships. This is more likely where open, informal dialogue is enabled, where the 'right' composition of people are in the room, and in contexts where existing networks are weak or dysfunctional but there is a positive wider discourse in support of EIPM. Where participants have the motivation or opportunity to utilise new relationships, they can be used to share information or advice, or can lead to new organisational collaborations	Not prioritised for investigation

CIMO 5. Foot in the door Providing individual-level support (such as training or coaching) in a sensitive, collaborative way can provide a 'foot in the door' for BCURE partners, generating permission and buy-in for them to begin implementing organisational reforms – this could be a particularly important 'way in' in contexts where it is not possible to start working directly at organisational level; for example, where access to government is difficult to secure	Tested
CIMO 6. Filtering up When a sufficient number of individuals (including some with leadership roles) begin accessing, appraising and applying evidence more in their work, this can 'filter up' and lead to higher-level recognition of the value of an evidence-informed approach – through senior staff seeing and being impressed by good-quality evidence products and through these products feeding into senior decision-making processes and improving them	Tested
CIMO 7. Cascading When individual support influences individuals in mid-level roles, who are committed and passionate and who have supportive senior management, they can formally cascade their learning through introducing new ways or working and new structures and processes within their organisations	Tested
CIMO 8. Showcasing Providing technical support to co-produce tools or systems that facilitate staff to use evidence more effectively, where this is done in a collaborative and innovative way, can generate good examples that 'showcase' the value of evidence for quality, performance and delivery. These 'showcases' provide user-friendly decision support tools that help individuals use evidence, but also build understanding and buy-in among senior staff about the value of evidence for decision making, resulting in examples 'diffusing' out to inspire new reforms elsewhere	Tested
CIMO 9. Accompaniment Where there is pressure to improve performance from senior levels and where an external partner has established trust through previous activities, this can enable an 'accompaniment' mechanism: high-level stakeholders give partners the permission to provide ongoing, tailored support to help them embed EIPM. This can lead to uptake of recommendations from processes facilitated by the partner, adoption of tools or systems, and possibly the emergence of an internal unit to 'own' and 'champion' EIPM	Tested
CIMO 10. Adoption Providing technical support to co-produce tools or systems that facilitate staff to use evidence more effectively can spark a high-level decision to formally adopt the tools or systems to help standardise EIPM within the organisation. This is more likely when they link to other government procedures and are backed by sufficient authority. Adoption can be on a small scale (e.g. adopting templates), but, in a context where there are high-level government 'owners' of EIPM, adoption can also be large scale (e.g. adopting a comprehensive policy and planning system to promote, embed and monitor the quality of evidence use throughout the policy cycle and into the future)	Tested
CIMO 11. Facilitation Tools or systems to promote EIPM sparking a facilitation mechanism – providing practical assistance enabling people to do their jobs better / more easily. This results in the system or tool being used, and (potentially) increasing the value of evidence through demonstrating the benefits it can bring	Tested
CIMO 12. Reinforcement Tools or systems that involve positive or negative incentives to adopt EIPM behaviours sparking a *reinforcement* mechanism, in which positive incentives or risk of negative consequences influence behaviour, and lead to individuals deciding to change the way they access, appraise or apply evidence in decision making	Tested
CIMO 13. Sustainable national hub Supporting local organisations to deliver EIPM capacity-building activities (directly through organisational capacity support, and/or indirectly through providing opportunities for national partners to 'learn on-the-job'), can strengthen organisational capabilities through 'learning by doing.' This can result in the establishment or strengthening of national institutional actors, which can act as a 'hub' for EIPM, are capable of running successful programmes to promote it and are potentially able to continue supporting it once the programme has ended	Tested only in Zimbabwe, which was the only context that aimed to build a sustainable national partner
CIMO 14. Relating and attracting Where local organisations successfully deliver programme activities and/or explicitly aim to build relationships with government departments and other EIPM actors, this enables partners to 'relate and attract' – providing exposure to new collaborators. This can lead to increased demand for partners to provide capacity-building support for EIPM from new actors not originally targeted by the programme – which can provide a crucial entry point where there are sensitivities around influencing government decisions, and hence where it is difficult for 'outsiders' to gain entry to government organisations	

Stage 3 final report: identifying impact pathways towards improved use of evidence

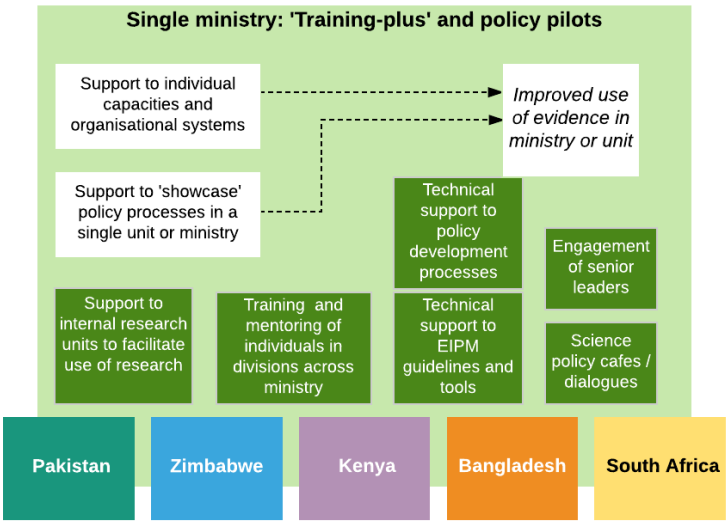
The Stage 3 synthesis process compared the contextualised country case study ToCs to identify how the tested CIMOs had played out in the different countries. This highlighted three main 'routes' towards EIPM taken by BCURE partners, at different levels of the government. They build on the idea of 'entry points at different levels' articulated earlier in the evaluation, but structure the findings in a more holistic, case-based way rather than disaggregating our findings according to levels of change. We have termed these three routes towards EIPM 'impact pathways':

- **Impact pathway 1:** Support to a single ministry or unit
- **Impact pathway 2:** Working at a government-wide scale
- **Impact pathway 3:** Support to parliament

The impact pathways are archetypal programme theories (Funnell and Rogers, 2011), presenting a sequence of activities and outcomes from short term to long term, with evidenced causal explanations of how and why change has come about through BCURE. They are not mutually exclusive – most projects worked across two or more. The three impact pathways take the place of an overarching, portfolio-level ToC or programme theory, providing a rich and context-specific explanation of how and why capacity support can promote EIPM through entry points at different levels.

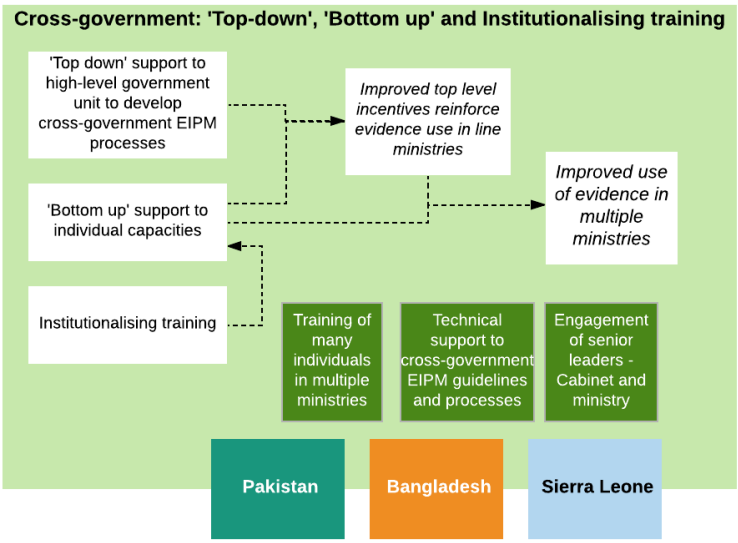
Impact Pathway 1: Support to a single ministry or unit

The single ministry pathway incorporates two approaches: 'training-plus', and technical support to pilot policy processes or EIPM tools. In the 'training-plus' approach, training on EIPM was given to technical officers responsible for policy formulation, who were then provided with follow-up support, to help sustain the application of new EIPM skills. Organisational support was also given to tools and guidelines that were intended to be adopted by ministries in order to facilitate and in some cases, incentivise and reinforce individuals to use evidence more routinely and more skilfully. In the second approach, some BCURE projects provided technical support at an organisational level to accompany pilot policy processes, in order to 'model' systematic, evidence-informed approaches to policy development within the ministry, provide EIPM trainees with opportunities to apply their skills, and produce evidence-informed policy proposals. Other projects provided technical support to the development of data and evidence tools that aimed to showcase the value of evidence for decision making, intending for them to be adopted or replicated by government partners to help facilitate and embed evidence use in the ministry or unit.



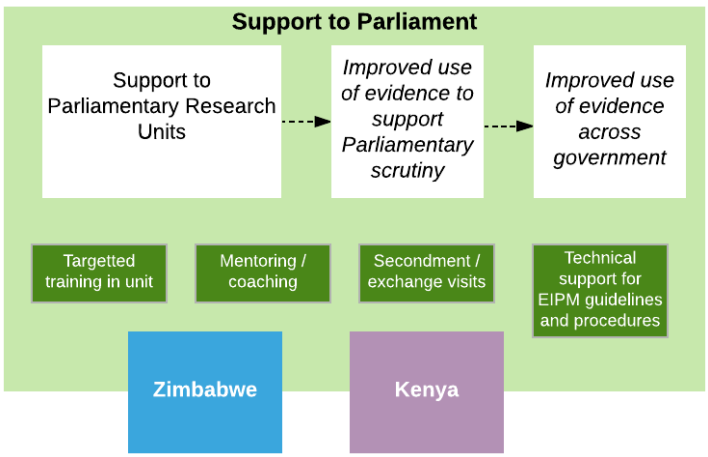
Impact pathway 2: Cross-government approach

The cross-government pathway incorporates three approaches: 'top down', 'bottom up' and 'institutionalising training' to promote EIPM. In the 'top down' approaches, BCURE supported activities such as working with cabinet to develop and roll out EIPM guidelines and procedures, often aiming to establish common cross-government standards for EIPM, facilitating various learning events, exchanges and training with senior government stakeholders to build high-level buy-in for EIPM, and supporting (mainly senior) officials in line ministries to develop evidence-informed policy processes, systems and structures in their ministries. In the 'bottom up' approach, BCURE programmes developed EIPM training courses and delivered it to large numbers of civil servants. The third approach involved BCURE also working to institutionalise EIPM training embedding EIPM curricula within national public sector training institutes. In all three projects working across government, at least two of the three approaches were adopted simultaneously.



Impact pathway 3: Support to parliament

This setting poses a different set of issues and challenges – as parliaments do not make policy, but can play an important role in interrogating it and holding line ministries to account. This pathway incorporates training research staff within a parliamentary research unit (as an entry point to parliaments), combined with follow-up individual and organisational support to strengthen parliamentary use of evidence in oversight and scrutiny functions. The two BCURE projects taking this approach promoted EIPM by focusing on individual and organisational change in parallel, as in the single ministry pathway. Training on EIPM was provided to a cohort, or all, research officers responsible for supporting MPs and committees with impartial and accurate analysis of policies and bills, and with evidence for committee enquiries. Trainees were then provided with follow-up support to help sustain the application of new EIPM skills, as well as offered learning exchange opportunities with other parliaments to further enhance learning about EIPM. BCURE also provided flexible, ongoing technical support, in collaboration with senior managers, at the level of the research unit to support trained researchers to cascade skills to non-trainees and develop procedures, ways of working, tools and guidelines that were intended to be adopted by parliament in order to facilitate researchers, and sometimes MPs themselves, to use evidence more routinely and more skilfully.



Testing and revising CIMOs

Within the overarching impact pathway frame, evidence on how and why changes had and had not happened were synthesised across the cases (method described in Section 3.4). This allowed us to formulate 'empirical CIMOs,' which unpacked 'what worked, for whom, and why' within each impact pathway in terms of the **outcomes** (O) that had come about in different case study contexts through various **mechanisms** (M), and the **context** (C) and **intervention** (I) factors that enabled or constrained the mechanisms. See below for an example.

Establishing 'top down' cross-government tools and structures: What worked, for whom, and why?

In **Bangladesh** and **Sierra Leone**, Cabinet Division and the Cabinet Secretariat decided to adopt and endorse new EIPM tools and systems (M, O) because these institutions had clear ownership over and buy-in to the process (C), in part a consequence of the support of high-level champions (C), and in part because of they had a mandate for reform aligned with BCURE's objectives (C). In **Bangladesh**, the framing of EIPM as a technical approach to improve policy formulation was a key selling point (I). Ownership was also promoted through BCURE's implementation approach, which was sensitive, flexible, and tailored to the local context (I): an approach that can be characterised as 'accompaniment' (M).

Following a high-level directive from Cabinet Division / Secretariat (C), EIPM tools and guidelines have been adopted by line ministries in **Sierra Leone**, and seem likely to be adopted in **Bangladesh** (M), where the policy pilots have to some extent successfully showcased (M) their value. However, insights from **Sierra Leone** and **South Africa** suggest that a one-off directive is not enough: ongoing engagement through 'carrots and sticks' (C) is necessary to ensure new tools are actually used (O).

These empirical formulations across the three impact pathways were then brought up to the level of middle-range theory, again following the synthesis approach described in Annex 3.4. The synthesis identified six key mechanisms that, when catalysed, led to positive changes around the use of evidence, although not all of them were always present in any one project. The key mechanisms do not operate in insolation, but instead work together to catalyse change, and build on each other so that where one mechanism operates it often creates a conductive context for another mechanism to 'fire'. These mechanisms are derived from well-established theories from psychology, sociology, development studies and governance – referenced throughout the report.

- **Accompaniment:** where an external partner provides tailored, flexible and responsive support to a government institution through a process of reform, characterised by a high level of trust, as opposed to a more traditional supplier / consumer model where ad hoc support is provided through one-off interventions. This often involves co-producing tools, systems or policy products.
- **Self-efficacy:** where providing information, opportunities to practise skills, coaching or technical support builds individuals' confidence in their ability to do their jobs or achieve a particular goal. This is akin to feeling of 'now I know how... (to find the evidence I need, to weigh up sources, to communicate evidence effectively).'
- **Facilitation:** where a tool, system or process for EIPM facilitates government officials to do their jobs or undertake a task more easily or efficiently.
- **Reinforcement**: where rewards or other forms of control create incentives that motivate officials to work in a particular way. Positive reinforcement includes rewards and encouragement, while negative reinforcement includes reminders, audits and mandatory requirements.
- **Showcasing**: where providing good examples of evidence tools or processes demonstrates the value of an evidence-informed approach, which leads to them being adopted elsewhere.
- **Adoption**: where senior government stakeholders decide to adopt a new tool, system or process for EIPM to help standardise EIPM within a government institution. This can be on a small scale (a unit

deciding to adopt a new template to standardise policy briefs) or a large scale (a government deciding to adopt a revised procedure for policymaking across all its line ministries that requires engagement with evidence). Adoption can happen for many reasons, and there is a risk that it might be a case of 'isomorphic mimicry' – where a new tool or system is adopted on the surface in order to access donor resources, without actually changing day-to-day practice.

The evaluation also identified a further mechanism that implicitly underpinned several BCURE project theories, but which has not (yet) catalysed in practice:

- **Critical mass**: where changes in practice among a sufficient number of government officials diffuse out to influence colleagues' behaviour, and the rate of adoption of new behaviours becomes self-sustaining. This diffusion may happen through cascading, where government officials formally cascade their new knowledge on EIPM through introducing new ways of working or new structures and processes. Or it may be through filtering out or filtering up: where improvements in evidence use by government officials leads to recognition of the value of an evidence-informed approach among colleagues (filtering out) or senior management (filtering up) which in turn influences' colleagues behaviour, or increases senior-level support for evidence-informed ways of working and/or organisational reforms to promote EIPM.

The final evaluation found that the key mechanisms do not operate in insolation, but instead work together to catalyse change, and build on each other so that where one mechanism operates it often creates a conductive context for another mechanism to 'fire'. Our revised CIMOs reflect this, representing our final tested theory about what works to build capacity for EIPM, for whom, and in what circumstances.

CIMO 1. Where there is genuine interest in partnership from high-level government stakeholders, existing incentives for evidence use in policymaking, and a window of opportunity to catalyse reform (C), an external partner can accompany EIPM reforms (M) in a participatory and collaborative way, providing tailored, flexible and responsive ongoing support that evolves over time (I) in response to emerging challenges and opportunities (C). This mode of working is greatly helped by the presence of high-level, enthusiastic and committed champions (C), and can create a conducive context for the other EIPM mechanisms to operate through encouraging government ownership (O) and building trust in the partner to work alongside government (O).

CIMO 2. Where information is provided about the importance of EIPM and how to access, appraise and apply evidence, alongside opportunities to practise skills, this can generate self-efficacy (M) and lead to individual behaviour change (O). Behaviour change is more likely to be sustained where there are clear incentives that motivate participants to apply their learning and reinforce changes in practice (M) – this includes management support to encourage and provide space for participants to access, appraise and apply evidence (C). Behaviour change is also more likely where activities are closely targeted to individuals who can apply their learning because it is directly relevant to their day-to-day work (I), and where activities are practical and participatory (I), provide practical tools (I) that facilitate trainees to do their jobs more easily (M), incorporate a focus on soft skills as well as technical skills (I), use knowledgeable, patient and confident facilitators (I), and tap into incentives to encourage participation (I).

CIMO 3. Where a cohort of officials start accessing, appraising and applying evidence more effectively, this can diffuse out to influence colleagues' behaviour (O) through a 'critical mass' effect (M). This is more likely when the cohort consists of a good number (I) of well-connected and clustered officials (C) in a unit with some reach and influence within the broader organisation (C), and where there are clear organisational incentives to use evidence (C) and senior management support to cascade learning (C) – potentially supported by a 'training of trainers' strategy (I).
Tentative theory, based largely on insights about blocking factors from Kenya and Zimbabwe, and insights from the wider literature.

CIMO 4. Where technical input is provided to support a policy process in an evidence-informed way, or develop a tool to improve evidence access, appraisal or use, this can generate high quality policies or

products (O) that showcase the value of evidence for quality, performance and delivery (M) and lead to adoption (O) and diffusion (O) of the procedure or tool. This is more likely where external actors 'accompany' government partners to co-produce policies or tools in a flexible, responsive and collaborative way (I), where policies are high priority or tools address a recognised problem (C), and where tools are intuitive and interactive (I) and genuinely facilitate officials to make decisions and do their jobs better and more easily (M). However, adoption can be stymied by many factors including shifting political priorities or resource constraints (see CIMO 5).
Strong evidence in support of theory, from Kenya, Pakistan, South Africa and Bangladesh (see single ministry pathway).

CIMO 5a. Where capacity support succeeds in showcasing the value of an evidence-informed approach, training course, tool or product (M) and /or generating tools that facilitate staff to do their jobs more easily (M), this can lead to a high-level decision to formally adopt the initiative to help standardise EIPM (O). Meaningful adoption is more likely where reforms have been co-produced by government and external partners through a flexible and collaborative process of accompaniment (C), and where there are high-level institutional and individual champions with a clear mandate for and interest in reform (C) who have access to resources to scale up or roll out the initiative (C). Adopted tools and processes, when attached to high-level incentives and encouraged through ongoing support rather than just a one-off directive (C), can then help reinforce (M) changes in practice at an individual and organisational level (O).
Strong evidence in support of theory from Bangladesh, Sierra Leone, Kenya, South Africa and Pakistan (see single ministry and cross-government impact pathways). Insights on factors that blocked adoption in Zimbabwe also support this theory.

CIMO 5b. Where capacity support succeeds in catalysing high-level ownership and buy-in to EIPM (C), it can position a national governmental unit to carry on promoting EIPM into the future (O), provided it is able to access resources (C) and buy-in is not eroded by high-level changes in priorities or staffing (C).
Tentative theory based largely on insights from the South Africa impact case study, and early observations in Bangladesh

The table below describes how the Stage 2 CIMOs were accepted or rejected, and how they translate into the final CIMOs at Stage 3.

Stage 2 CIMOs	Status at Stage 3	Corresponding Stage 3 CIMOs and supporting evidence
CIMO 1. Self-efficacy	This theory was not directly investigated at Stage 3, as there was strong evidence to support it at Stage 2, and the final evaluation prioritised theories relating to organisational change. However, several additional insights were gained through Stage 3 relating to the contextual and implementation factors that help this operate, including about how this mechanism interrelates with 'facilitation' and 'reinforcement' to help lead to routine behaviour change	CIMO 2
CIMO 2. Coaching	Not prioritised for investigation	N/A
CIMO 3. Learning from similar challenges	Not prioritised for investigation	N/A
CIMO 4. Facilitated	Not prioritised for investigation	N/A

Stage 2 CIMOs	Status at Stage 3	Corresponding Stage 3 CIMOs and supporting evidence
spaces for dialogue		
CIMO 5. Foot in the door	Tested: rejected. The Stage 3 evaluation unpacked a much more nuanced explanation of why BCURE partners gained a 'foot in the door', linking to 'thinking and acting politically' and accompaniment. These factors proved significantly more important in providing a foot in the door than starting the programme through training, across all four Stage 3 cases.	N/A
CIMO 6. Filtering up	Tested: Very limited evidence that this theory held; however insights on blocking factors from Kenya and Zimbabwe	CIMO 3. These theories are both facets of 'critical mass,' a theory implicitly underpinning several BCURE projects. Although there is only tentative evidence for them, there are insights from some contexts and the wider literature on blocking factors that inhibit the theory from holding.
CIMO 7. Cascading	Tested: Some evidence, at a relatively small scale, in the Kenyan MoH and parliament, and insights on blocking factors from Kenya and Zimbabwe	
CIMO 8. Showcasing	Tested: confirmed and nuanced with additional insights from Stage 3.	CIMO 4. Strong evidence in support of theory from Kenya, Pakistan, South Africa and Bangladesh
CIMO 9. Accompaniment	Tested: confirmed and nuanced with additional insights from Stage 3.	CIMO 1. Strong evidence in support of theory from across all six settings
CIMO 10. Adoption	Tested: confirmed and nuanced with additional insights from Stage 3.	CIMO 5. Supported by strong evidence in support of theory from Bangladesh, Sierra Leone, Kenya and Pakistan. Insights on factors that blocked adoption in Zimbabwe also support this theory
CIMO 11. Facilitation	Tested: confirmed and nuanced with additional insights from Stage 3.	Rather than standing alone, this mechanism was an important part of the context that helps spark routine behaviour change (CIMO 2), showcasing (CIMO 4), and adoption (CIMO 5). Supported by insights from Pakistan, South Africa, Kenya, Sierra Leone and Bangladesh
CIMO 12. Reinforcement	Tested: confirmed and nuanced with additional insights from Stage 3.	Rather than standing alone, this mechanism was an important part of the context that helped spark routine behaviour change (CIMO 2) in Kenya, Zimbabwe and Sierra Leone, was deemed important to incentivise behaviour change in Bangladesh, and was one of the main reasons for many trainees *not* changing their behaviour in Pakistan. This mechanism can also spark as an outcome of adoption (CIMO 5)
CIMO 13. Sustainable national hub	Tested in Zimbabwe, which was the only context that aimed to build a sustainable national partner. Theory was broadly confirmed, with some additional insights captured, explored in the VakaYiko case study report. However, this is not captured in the Stage 3 CIMOs given the limited evidence base or applicability to the other projects	N/A
CIMO 14. Relating and attracting		

5 Political economy analysis template for Stage 3

This template was provided to national consultants to complete prior to primary data collection

PEA purpose: Mapping of country and sector-level institutional arrangements

PEA focus: [Country] [Ministry] [Cabinet] [Parliament]

Purpose: To produce a descriptive mapping of national and sector-level political dynamics that affect policymaking in the focus sector.

PEA focus: Country – Specific Sectors- e.g. Ministry of Health and Parliament

1. **Objective:** To produce a **descriptive map** of the institutional and political dynamics around policymaking and parliamentary scrutiny of policy decisions. This should use as a starting point the contextual summary provided in the Stage 2 programme evaluation report.

2. **Process:** Please respond to the questions on Part 1 and 2 of the question framework below. Use the headings in bold to structure the document. This is an internal document, so please only provide informative bullet points or short sentences to provide information for each heading and sub-bullet. No need for a polished narrative.

   Also, please add any additional information that you feel is important but doesn't fit neatly under any headings.

3. **Input and output**: The document should be between 3-6 pages long. You should spend no more than 3-4 days on it.

4. **Data sources:** The consultant will need to use secondary sources, and possibly a key informant interview with the BCURE team.

Secondary sources need to be selected with consideration as to their relevance, reliability, accuracy, independence, timeframe and sourcing of the information. Wherever possible attempts should be made to corroborate the information used across independent sources, to ensure accuracy.

All sources cited need to be referenced in footnotes, with weblinks, following the Harvard style (see https://www.imperial.ac.uk/media/imperial-college/administration-and-support-services/library/public/harvard.pdf), and listed at the end of the document.

Secondary sources could include the following:

- Information produced by international sources, e.g. international non-governmental organisations country reports on health sector
- Information produced by bilateral and multilateral organisations, e.g. WHO,; World Bank; and/or United Nations Organisations' country reports; USAID, other donors, e.g. https://www.healthpolicyproject.com/index.cfm?id=country-kenya
- Information produced by other governments, for example country analyses produced by the UK government
- Independent media reports, e.g. *Financial Times*
- Information produced by the SECURE programme
- Information produced by the Kenyan government on the functioning of the Ministry of Health, and the various acts it is implementing

A key informant interview with the BCURE team should be conducted to gather sector-specific information that is not in key documents, and to enquire about additional sources and possible interviewees.

Identifying key informants

Some of the below questions will need to be explored through interviews, as the information will not be available through secondary sources. Through speaking to the BCURE partner, please identify 6–7 individuals who might be able to give insights into the questions below. (Some may be informants who have knowledge across the government system, while others have sector-specific knowledge relating to the Ministry of Youth / parliament. These may include civil society stakeholders, key political analysts, academics and government political reporters.)

PEA topic areas and questions

Please provide informative bullet points, no need for a polished narrative.

PART 1: General update on national-level issues affecting the [sector] since June 2016
To be completed before the country visit starts.

Overall events and trends in the [country] context since June 2016:

- What regional, national or international events / issues are having a major influence on life in [country]?
- What major issues and events have affected government and policymaking? E.g. economic, political, social, socioeconomic, environmental, health etc.?
- Any new government-wide initiatives introduced, e.g. anti-corruption measures; regulations; transparency?
- Have political incentives, ideologies or values changed and how is this affecting behaviours of politicians and citizens?
- What is the influence on today of historical legacies and change processes?

PART 2: Specific situation at a sectoral level (focusing on 2013–16). To be completed as far as possible before the country visit starts.

[Ministry]	[Parliament]
Actors and key players: *The actors (individuals or collectives) involved in making, influencing and delivering policy, including actors at different levels of government including those at sub-national and regional level*	
• What is the official status, role and the mandate of the Ministry? • How is the ministry structured and what are the key departments? • Who have been the dominant individuals within the Ministry, what is their role? • Have there been any reorganisations within the Ministry of since 2014? • Where does the **[xxx unit]** fit in? What services does it provide? • Who are the other key players that have an influence on youth policymaking, implementation and priority-setting, beyond the ministry? *Please include both national, county and international* • Where does the Ministry of Youth sit in relation to national and county-level policymaking and priority sectors? • How important and/or influential is youth as a sector in national politics?	• What is the official status, the role and the mandate of the parliament? • How is parliament structured and what have been the key committees (select committees; standing committees etc.) from 2013-16)? • Who have been the dominant individuals within parliament, what is their role? • Where does the **[xxx unit]** fit in? What services does it provide? • How much influence does parliament have in government and public life, and what is the basis for this? • What power do other actors have to scrutinise parliament?

[Ministry]	[Parliament]
• How is the Ministry portfolio viewed by political appointments, and why? E.g. desirable, problematic, lucrative?	
Institutions: *The rules, norms, practices, relationships which influence individual and collective behaviour. These may be formal or informal, and exist at different levels within governments and departments*	
• Are the rules, roles and responsibilities for youth policymaking formally set out in the constitution? • Which entity at which level (national, sub-national etc.) is responsible for leading the policymaking process? • What changes have there been in the balance of power between national, county-level and international youth actors between 2013 and 2017? What are the implications of these for youth policymaking? • How are finances and resources allocated in Min of Youth, e.g. which actor approves budgets, e.g. cabinet and/or parliament? • Which key players provide the financing, e.g. international donors?	• What is the role played by parliament in national policymaking and priority sectors? (Its position from 2013-16, e.g. minor sector or major player in cabinet) • What is parliament's role in budgeting process and procedures, who approves budgets? • What are the formal and informal rules for scrutiny of policymaking and implementation between parliament and the executive? • What are the main Parliamentary bodies that provide scrutiny over the Ministry of Youth?
Policy networks: *the relationships between actors responsible for policy decisions and those who seek to influence it; the level of influence these groups have and the level of government consultation with them*	
• Have there been changes in the players/actors and/or interest groups that have influence on youth policymaking now? • What are the interest groups that the [Ministry] responds to? E.g. private sector health providers; international donors etc.? • What is the role of external actors on government policymaking (e.g. international donors; lobby groups; civil society groups)? • Which groups does the government consult most, and which coalitions seem to have the most influence (both in the youth sector and more broadly?) • What are the mechanisms for consultation with citizens? How are citizens are involved in policy development and monitoring (e.g. referendum, opinion surveys)?	• What inter-linkages are there between parliament and other organisations inside and outside government, e.g. policy analysts; universities; parliamentary committees, others? • What is the role of external actors on policy analysis and scrutiny (e.g. international donors; lobby groups; civil society groups)? • What are the mechanisms for consultation, participation and inclusion in policy processes and the way in which citizens are involved in policy development and monitoring (e.g. referendum, opinion surveys)?
Context: *The socioeconomic, demographic, and geographic factors that policymakers take into account when making decisions, and the routine (e.g. elections) and non-routine events with the potential to shift attention or provide an impetus to policy change*	
• Elections 2018: How is youth being presented and talked about in the media and social media as an issue in Zimbabwe in the run up to the elections? • What have been the events /issues that have affected specifically **[sector] policymaking** and **service delivery** in [country], e.g. scandals, strikes, protests etc. How has the government responded to events, negative and positive?	• Elections 2018: How is the run up to elections influencing parliament? • What have been the events /issues that have affected parliament specifically in [country], e.g. scandals, strikes, protests etc.?
Ideas: *ways of thinking and the extent to which they are shared within groups, organisations, networks and political systems. This includes the interplay between different forms of knowledge underpinning*	

[Ministry]	[Parliament]
action, the often deeply held beliefs of actors, and the ability of actors to persuade others to act in a particular way (e.g. through framing issues in specific ways)	
• What have been the main **[sector] policy priorities** between 2013 and 2016? • What are the main citizens' concerns about youth? • What are the narratives being presented in the media and social media about youth policies and service delivery? • Have there been changes in the beliefs, ideologies and values which shape the youth sector?	• What are the predominant values, narratives and perceptions that influence policy discussions in parliament, and what is the source of these narratives?

PART 3: Questions to be integrated into interview topic guides

Leadership, management and organisational culture:

- What are the leadership and management structures in the wider [Ministry], and what is their basis?
- What are the incentives and motivations that influence staff? Formal and informal, positive and negative?
- How do resource, capacity and skill levels vary across the organisation, including among managers and leaders, and with what consequences?
- How hierarchical is the organisational culture? To what extent can technical staff and political appointments challenge peers and seniors and/or express alternative views on policy issues?

Institutions and rules around policymaking:

- What is the influence of leadership, management on the content and direction of policy?
- In what specific ways does the **[xxx unit]** contribute to policy processes?
- What is the influence of dominant or prominent personalities on policymaking processes?
- What are the predominant values, narratives and perceptions that influence policy formulation, and what is the source of these narratives?
- How do power relations influence policy negotiation processes?
- What role does evidence, data and evaluation play in policy and decision making?

6 Sampling guidance

The below guidance was provided to BCURE lead evaluators prior to the Stage 3 country case studies

BCURE Stage 3 sampling guidance

Core principles of sampling

The sampling for Stage 3 is purposive, guided by the priority outcomes selected within the country-level programme theory. The aim is to achieve a *sufficient degree of confidence* in our hypotheses about the extent to which priority outcomes have occurred (EQ 1), BCURE's contribution to the outcomes (EQ 2) and how and why BCURE contributed or failed to contribute (EQ 3).

Our Stage 3 sampling approach is guided by four main principles. Bear these in mind and try to keep them front-and-centre when developing and iteratively revising the sample!

1. **Sampling will be driven by theory.** The starting point is priority outcomes and evidence tables. Who do we need to speak to in order to generate evidence for hypotheses?

2. **Sampling will be iterative,** allowing for changes and additions during field work as theories develop and leads are followed. The sample will therefore continually evolve throughout the data collection process.

3. **Sampling will aid comparison between sub-groups:** A key element of our sampling strategy will be *comparison* between different sub-groups of participants, in order to investigate how change has or has not occurred for different people / units / ministries etc., and to help explain how and why these differences exist. Sub-groups will evolve over time as our understanding develops. Although we will be limited by resources as to the number of sub-groups it will be possible to explore, we will be guided by the emerging evidence on what seems to be most important in explaining the outcomes.

4. **Sampling will seek to maximise triangulation of sources for each hypothesis**: We will aim to triangulate evidence against our hypotheses across a range of different stakeholders, through comparing insights from project participants with insights from knowledgeable 'outsiders' (informed by the PEA of who is influential in relation to the outcome), and through accessing secondary documentation where available. Our data sources are detailed below.

Five categories of data sources:

The aim is to triangulate insights for each priority outcome from as many of the below categories as possible.

1. **Monitoring data and other programme documentation**. This will be reviewed first to examine secondary evidence for hypotheses. It will also help identify relevant sub-groups of individuals to target for interviews.
2. **Interviews and workshops with programme staff**. This will supplement the monitoring data, helping understand gaps or areas where greater clarity is needed. It will also provide an insight into the areas project staff think have been more or less successful and how and why, which will help further develop our theories.

3. **Interviews with project participants (individuals directly targeted by the activity / activities which aimed to contribute to the outcome)**. This will generate self-reported insights about the extent to which outcomes have been achieved or not achieved, how and why, for different groups.

4. **Interviews with other knowledgeable stakeholders.** These are stakeholders who did not participate in BCURE interventions, but who can provide insights into (a) outcomes observed and the reasons for these outcomes; and (b) political economy issues that relate to priority outcomes. This group will be considerably enlarged this year, in order to address concerns of the Steering Committee and EQUALS review that the Stage 2 report did not sufficiently deal with potential confirmation bias from project participants.

5. **Secondary (non-project) documentation.** This is documentation *not* produced by the programme, which provides insights into outcomes and the reasons for outcomes. In previous years, it has not proved possible to access this documentation, and there remains a major risk that important documents will remain impossible to access this year. However, in previous years we have been unable to dedicate much time to identifying and attempting to secure relevant documents, rather relying on BCURE programmes to share documents they had access to. This year, time will be dedicated to identifying and securing potentially relevant documentation (a) up-front when evidence is assembled; and (b) throughout the data collection phase, using interviews to attempt to secure documents that can triangulate insights from respondents.

Sampling rules of thumb

- Aim for roughly **60** interviews in total. However, don't feel the need to interview people for the sake of it! If it is not possible to reach this many people with knowledge of the programme, you might want to do additional interviews with PEA informants, or focus more resources on trying to get hold of (and then reviewing) secondary documentation.

- Around **30–35** interviews with *programme participants* (predominantly government, but also civil society and non-government stakeholders where they have been direct targets of the programme).

    - Consider the 'rule of three': where possible, aim to speak to 3 people from each relevant 'for whom' 'sub-group' – see below

- Around **20** interviews with people who were not involved in the programme but who can give insights into whether outcomes were achieved / how and why, and PEA factors.

    - This should include around **5–7** interviews with people who may have no knowledge of the programme but who can give insights into political economy factors relating to the relevant sectors (e.g. political analysts, academics, think tank or civil society stakeholders, DFID staff)

- You should conduct an in-country workshop (and if you think relevant, individual interviews) with BCURE staff.

- If relevant to understanding activities conducted since last year it may be helpful to conduct a small number of interviews with trainers, mentors, and other facilitators of activities.

- Interviews don't all need to be lined up in advance – there should be flexibility to add new stakeholders once in-country.

- Sampling decisions should be transparent: documented clearly using the sampling spreadsheet. The sampling spreadsheet should also be used to identify in advance which outcomes and CIMOs will be tested with which people, to ensure we are being systematic.

Sub-groups

There are two types of 'sub-group' we are interested in:

1. Different sub-groups **targeted by the programme**. These may be individuals from different target ministries, units or departments, male and female participants, more junior and more senior participants.

2. Sub-groups **associated with differential outcomes**. These may or may not be the same as the sub-groups targeted by the programme. This is essential for testing our theories (CIMOs) about how and why BCURE works, and understanding *for whom* BCURE works. For example, did some senior staff demonstrate more buy-in to an EIPM agenda following an intervention than others? If so, it will be important to try to speak to individuals (and if possible their colleagues / peers) from the 'more buy-in' and 'less buy-in' sub-groups in order to understand what it was about these individuals or the wider context that enabled or constrained buy-in. Iteration is vital – as our theories and our

understanding of differential outcomes evolve over the course of data collection, new sub-groups will emerge and others disappear or subsume into broader groups.

Steps in the sampling approach

1. Start with last year's sample. Either save a new version of last year's spreadsheet and amend it to reflect the Stage 3 sample template (saved in Dropbox), or copy relevant stakeholders into the Stage 3 template. Please do make sure you're using the Stage 3 rather than Stage 2 template, as changes have been made to the stakeholder categories etc.

2. Work through each priority outcome in turn and consider who will be able to give insights into the EQ 1, 2 and 3 hypotheses, and then add them to the sample. You might know these people by name from last year's stakeholder lists, or you might just indicate their role at this stage [e.g. 'someone high level in the MoY']. Note that you don't need to do this separately for outcomes 17a-d, and impact level change – this should be covered through considering all of the previous outcomes.

    i) Make sure you've included both participants in BCURE activities and non-participants, for each priority outcome. Non-participants might include:

        - Managers or colleagues of participants
        - Senior stakeholders from the department
        - Members of other units or teams who work with participants' teams

    ii) Make sure you've thought about potentially relevant sub-groups of participants in relation to that outcome, and where feasible tried to include 3 members of each.

        - E.g. three junior and three senior trainees from the Ministry of Youth

    iii) Start compiling a list of potentially useful secondary sources of evidence relating to that outcome, to try and track down.

3. Draw on the following sources in order to continue building the sample in line with the theory

    i) The evidence we already have for each of the priority outcome hypotheses (in the evidence tables). This should give insights into the additional data needed (on top of what we already have) to collect to test each part of the theory.

    ii) The document review, which should give more insights into who might be important to test particular outcomes, and which may also include participant lists to draw on when developing the sample.

    iii) The PEA review from the national consultant, which may give insights into people who can provide insights into various outcomes or into PEA issues.

4. Share draft sample with partner for comments, additional suggestions and contact details, and ideas about who should be a priority to talk to.

5. Once fieldwork is under way, ask interview respondents to identify further individuals who can provide an insight into a particular outcome, or who are members of a particular sub-group that is emerging as important. This strategy will be crucial to identify knowledgeable non-participants in BCURE interventions, who may be unknown to programme staff and therefore difficult to identify up-front. Also make sure you ask respondents about any potentially relevant documentary evidence.

Sampling spreadsheet

This spreadsheet was used by country case study leads, to record details of potentially relevant stakeholders and aid with the iterative and purposive sampling process.

	Stakeholder type	If programme participant - which interventions?	Priority outcomes to test / PEA stakeholder	CIMOs to test	Include/ exclude decision	Rationale for inclusion / exclusion	First name	Family name	Sex	Organisation	Job title	Government stakeholder?	Location	Interviewed at Stage 1?	Interviewed at Stage 2?	Email	Phone	Comments
1																		
2																		
3																		
4																		
5																		
6																		
7																		
8																		
9																		
10																		

7 Topic guides

7.1 Instructions for customising topic guides, for case study leads

Topic guides will need to be contextualised for individual stakeholders, around the outcomes and CIMOs prioritised for investigation. At Stage 3, we are speaking to three broad categories of stakeholders:

Programme participants	Individuals (government or non-government) directly participating in BCURE interventions (training, mentoring, workshops, knowledge cafes, policy dialogues, discussions around organisational systems development etc.)
Non-programme participants	Other knowledgeable stakeholders, who did not participate in BCURE interventions, but who can provide insights into a) outcomes observed and the reasons for these outcomes, and / or b) political economy issues that relate to priority outcomes
Programme staff	Individuals managing the programme, in-country and in the UK, including consortium partners. This also includes individuals hired by the BCURE partner to deliver training and mentoring, facilitate sessions etc.

- Prior to data collection, you will need to develop specific outcome probes for each of the priority outcomes in the country-level ToC. The aim is to find out what happened, and test the extent to which the outcome in the ToC actually emerged. The document review should help guide the formation of the questions (e.g. mentioning specific outcomes that you want to validate).
- As your understanding of the context and emerging outcomes develops, the outcome tables are likely to need refinement and new questions added to test emerging and more specific outcomes.
- We have developed CIMO tables to help probe specific priority CIMOs across the cases. These can largely be copied and pasted as-is into the topic guides, although the introductory questions and prompts may need some contextualisation. You'll need to decide which CIMOs are most relevant to which interviews in advance.
- You should decide in advance, as part of the sample development, which outcomes and which CIMOs to discuss with respondents.
- We have found that it is possible to probe a minimum of 2 outcomes and 2 CIMOs in depth within one interview (not including the longer-term outcomes 17a-d which have a separate set of questions at the end of the guide). Sometimes it is possible to test a much larger number of outcomes and CIMOs – but this is dependent on the flow of the interview and the extent to which change has or has not been observed (where change is minimal, it is easy to run through a larger number of outcomes).
- Use the sampling spreadsheet to keep track of which outcomes and which CIMOs you have discussed with which respondents and make adjustments if necessary, to ensure that you are testing the theory of change systematically.

Guide to developing probes for outcomes

EQ 1 probes

- Insert questions to examine the extent to which priority outcomes have come about.
- You should aim to ask a question for every link in the ToC that you want to test (e.g. if two arrows point to an outcome representing two specific causal pathways, and you want to test both links, you should aim to ask about them both).
- You will need to adjust and add new questions as the data collection progresses, to test emerging and more specific outcomes. For example if someone mentions a very specific outcome, it will be important to test this with others in order to triangulate insights.

EQ 2 probes

- Insert questions to examine the contribution of BCURE to priority outcomes.
- You should aim to ask 'why' or 'why not' for every outcome you are testing in the interview.
- Sometimes generic questions will be sufficient:
    - *What were the drivers and influences that led to this / prevented this from happening?*
    - *What do you think caused these changes / what is inhibiting change?*
    - *What was BCURE's contribution to these changes / initiatives?*
    - *Apart from the BCURE programme, has anything else contributed?*
- But sometimes it will be helpful to ask specific contribution questions relating to the outcome, especially for outcomes expected to emerge directly from BCURE activities. Other contributory factors may emerge as important during the data collection, which you may want to probe:
    - *Apart from BCURE, have you attended any other training courses / learning exchanges relating to evidence use?*
    - *Have you heard of xxx initiative? Do you think this contributed?*

Example outcome probes

Outcome 3. Government stakeholders apply, promote and communicate evidence routinely in their day-to-day work due to training…leading to… ***Outcome 8. Wider cohort of officials (beyond initial trainees) accessing, appraising and applying evidence more***
EQ 1 questions **[For programme participants]** • ***Individual.*** Has anything happened about how <u>you</u> work with evidence in your day-to-day work since the programme started? Can you give me some examples? Are there any written examples of work you can share with me? • What have you noticed about how <u>your colleagues</u> are working with evidence on a day-to-day basis – have there been any changes? Can you give me some examples? • ***Non-trainees in unit / dept*****:** Have you seen any signs that the training has influenced people who weren't actually trained? Can you give me some examples? • ***Senior staff:*** Have you seen any signs of the training influencing senior staff? Can you give me some examples? **[For non-participants]** • ***Non-trainees in unit / dept.*** Has anything happened about how <u>you</u> work with evidence in your day-to-day work since the programme started? Can you give me some examples? Are there any written examples of work you can share with me? • ***Trainees*****:** What have you noticed about how <u>your manager or your colleagues</u> <u>who went on the training</u> are working with evidence on a day-to-day basis – have there been any changes? Can you give me some examples? • ***Senior staff:*** Have you seen any signs of the training influencing senior staff? Can you give me some examples?
EQ 2 questions • What do you think caused these changes? • Apart from the BCURE programme, has anything else contributed? • What other training programmes or capacity-building opportunities are available within your unit / Ministry? Have you taken part? What was the content / what did you learn? • Are you involved in any donor programmes at the moment? Do you think this has contributed?

CIMO question tables

CIMO 5: Foot in the door	
One idea we have is that starting with a relatively neutral intervention like training might have provided a 'foot in the door' for BCURE. In other words starting with training generated permission and buy-in for them to begin implementing organisational reforms.	
IF YES • Why do you think this was? • What was it about the way BCURE engaged or provided the training that was important? Is it important to be collaborative? Is it important to be flexible? • Would BCURE have been able to come straight in and work at that level? What is it about the context that makes this type of approach important? (Probe from PEA framework)	**IF NO** • Why do you think that didn't happen? *(Probe for C and I factors that might have blocked the mechanism)*

CIMO 6: 'filtering up'	
One idea we have is that when enough people begin using evidence in a department, this can 'filter up' and make senior staff and peers recognise the value of an evidence-informed approach. Has it worked at all like that here?	
IF YES • Can you give examples? • What helped this 'filtering up' to happen here? • What other incentives were there, from the organisation or management? • How important was it that a group of people were trained at the same time? • How did peers/junior colleagues [who were trained] inspire / help you to work with evidence? • What other factors were at play? ○ *Need to probe for PE factors from the framework*	**IF NO** • Why do you think that didn't happen? *(Probe for C and I factors that might have blocked the mechanism)*

CIMO 7: 'Cascading'	
One idea we have is that when enough individual people, sometimes in mid-level roles, have been trained in using evidence, they can cascade new skills or introduce new ways of working with evidence to their teams. Has it worked at all like that here?	
IF YES • Can you give examples? • What helped this 'cascading' to happen here? • What other incentives were there, from the organisation or management? • How important was it that mid-level managers [who were trained] were committed and passionate about EIPM? • How important was it that they had senior management support? • What was it about how junior colleagues [who were trained] that helped them to inspire you to work with evidence? • What other factors were at play? ○ *Need to probe for PE factors from the framework*	**IF NO** • Why do you think that didn't happen?

CIMO 8: 'Showcasing'	
One idea we have is that tools or systems that help staff to use evidence more effectively (for example xxx) can act as what we call 'showcases' - good examples that promote and highlight the value of evidence. Has it worked at all like that here?	

IF YES
 	
• Can you give examples?
• What helped this 'showcase' to happen here?
• How important was it that the showcase used innovative approaches?
• How important was it that the tool provided opportunities to 'learn-by-doing' for staff?
• How important was the tool development was undertaken in a collaborative way by the partner?
• What other incentives were there, from the organisation or management?
• What factors helped the showcases inspire other reforms or new approaches?
• What other factors were at play?
  ○ *Need to probe for PE factors from the framework*
  ○ Did it matter which policy was chosen for support?
	IF NO
	
	• Why do you think that didn't happen?

○ Who decided?	

CIMO 9: 'Accompaniment'	
One idea we have is that when a government unit has developed trust in an external partner through a few collaborative activities, they allow the partner to 'accompany' policy processes and help embed evidence use. We call this 'accompaniment' - basically providing close-up, tailored and flexible technical support. Has it worked at all like that here?	
IF YES • Can you give examples? • What helped this 'accompaniment' to happen here? • How important was it that there was already a pressure to improve policy development processes from senior levels? • What other incentives were there, from the organisation or management? • What factors promoted trust in the partner? • Does this 'accompaniment' also help to optimise the government unit and strengthen their abilities to champion EIPM internally? In what ways? • What factors helped the internal unit to optimise its own work to promote EIPM internally? • What other factors were at play? ○ *Need to probe for PE factors from the framework* ○ Did it matter which policy was chosen for support? ○ Who decided?	**IF NO** • Why do you think that didn't happen?

CIMO 10: 'Adoption'	
One idea we have is that if an external partner provides technical support to co-produce tools and processes for using evidence, this can spark a high-level decision to formally adopt them as official procedures to help standardise and embed evidence use within the organisation. Has it worked at all like that here?	
IF YES • Can you give examples? • What helped this 'adoption' to happen here? • What other incentives were there, from the organisation or management? • How important was it that the tools/procedures were collaborative? • How important was it that they were backed by senior stakeholders/managers? How did that support come about? • How important was it that the EIPM procedures link to other formal processes? • What other factors were at play? ○ *Need to probe for PE factors from the framework*	**IF NO** • Why do you think that didn't happen?

CIMO 11: 'Facilitation'	
One idea we have is that evidence tools can provide practical assistance to people, essentially helping them do their jobs better or more easily, which means that tools actually get used. Has it worked at all like that here?	
IF YES • Can you give examples? • What helped this 'facilitation' to happen here? • How important was it that the tool helped you to do your job better? [benefit] • How did it do that, can you give me an example? • What other incentives were there, from the organisation or management to use the tool? • What other factors were at play? ○ *Need to probe for PE factors from the framework*	**IF NO** • Why do you think that didn't happen?

CIMO 12: 'Reinforcement'	
One idea we have is that when there are incentives to use evidence tools or procedures [e.g. xxx], this 'reinforces behaviour' – basically providing positive or negative incentives that lead individuals to change how they work with evidence. Has it worked at all like that here?	
IF YES • Can you give examples? • What helped this 'reinforcement' to happen here? • What other incentives were there, from the organisation or management? • How important was it that the procedures had senior management backing/authority? • How important was it that the procedures include monitoring use of evidence? • What other factors were at play? ○ *Need to probe for PE factors from the framework*	**IF NO** • Why do you think that didn't happen?

CIMO 13: 'Institutional local actor catalysed'	
One idea we have is that local organisation, delivering EIPM technical support as part of an international consortium, can lead to that local actor becoming optimised as a 'hub' for EIPM support nationally, beyond the end of the programme. Has it worked at all like that here?	
IF YES • Can you give examples? • What helped this to happen here? • Is it important that the actor is part of an international consortium? Does that help to build credibility / provide access? • How important was the 'learning by doing' aspect (through being part of an international consortium?) • What other incentives might have stimulated this, from the national sector or internationally? • What other factors were at play? ○ *Need to probe for PE factors from the framework*	**IF NO** • Why do you think that didn't happen?

7.2 Guide for workshop with programme implementing teams

1. Aims

- For the evaluator to understand fully what interventions have been implemented
- To check and validate the country ToC produced by the evaluator
- To explore EQs 1, 2 and 3 in relation to the 'priority outcomes' in the country ToC

The workshop will focus solely on the evaluation case study country, and will concentrate on the sectors / ministries / activities relating to the priority outcomes identified through the evidence mapping.

2. Set up and materials

- The session will be informal and participatory.
- With participants' consent, we would like to record the session to ensure we have an accurate record of the team's insights.
- It would be helpful to have access to a whiteboard, or a screen / wall where it is possible to put up post-its and flip chart paper.
- We will bring post-its, but if flipchart paper and marker pens are available this would be helpful. If not we can bring these with us.

3. Practical tips for facilitators

- Recording the conversation is a good back-up for detailed notes taken in the session.
- The workshop can be treated as a structured group discussion, but post-its and CIMOs on flipcharts are useful visual prompts.
- It may be worth preparing post-its or flip chart in advance, with lists of activities and outcomes drawn from the Stage 2 reports and the document review.

4. Agenda and process

Session	Details	Instructions for facilitator
Session 1: **Introduction** **10 mins**	Introduction Discuss the aim of the workshop and its role in the data collection Update on revised evaluation approach for Stage 3	**During introduction session:** • Explain purpose of the session – describe aims • Get consent to record • Introductions around the room • If you think relevant, potentially repeat the Stage 2 Icebreaker – card sort – how do you feel about the evidence-informed decision-making landscape in this country now that the programme has ended / is shortly ending?
Session 2: discussion of programme activities **30 minutes**	Discuss key programme activities in the case study country, additional activities since the Stage 2 evaluation, and rationale for any changes	Map out programme activities using post-its / flip chart. Begin with a list of activities from Stage 2 / drawn out from the document review, and verify these – using the opportunity to clarify any questions that may have arisen from the document review about what the programme has done and why *Check that there are no gaps in our understanding of what has been done.* *Clarify understanding of stakeholder groups – both department and role, and more conceptual (higher vs mid-level gov; colleagues from same dept vs people from diverse depts). Also probe for the programme team's rationale for targeting particular groups, through these*

Session	Details	Instructions for facilitator
		interventions. Why were the interventions combined and sequenced in this way? *How does gender come into your understanding of stakeholder groups?* *Looking back to the project plans this time last year – has there been any evolution in terms of your focus or activities?*
Session 3: Discussion of EQ 1 **45 minutes**	Discussion of EQ 1: **To what extent have priority outcomes been realised and for whom?** *a) Have the changes hypothesised in our country ToC happened?* *b) How does change differ for different sub-groups, organisations etc., reflecting on gender and equity issues?* *c) How sustainable is the change?* Also highlighting any examples of policy processes that the programme may have influenced, which may be the focus of the embedded policy case studies Throughout the discussion, highlighting relevant data sources (individuals to speak to and documents) that will help the evaluation to evidence outcomes and programme contribution.	Map out outcomes observed at Stage 2. Suggest preparing this in advance using post-its / flip chart. Put outcomes up alongside activities (starting to build the ToC) Ask: what else has happened since last year? Have any of these outcomes deepened, or led to further outcomes? It may be helpful to have pre-prepared post-its representing higher-level outcomes in our country ToC – to bring into this conversation in order to verify the ToC (is this how you saw things happening? Or was it something else?) Has anything *not happened* that you hoped would happen, or has anything taken longer or been more difficult than expected? (Again, post-its representing outcomes in the country ToC might help facilitate this discussion) Put up our priority outcomes 17a-d. Discuss whether the programme sees progress towards those outcomes, and where the links are from earlier outcomes. **Discuss sustainability:** • Looking at the changes observed so far, what do you hope / expect will happen now that / once the programme has ended? • What is it about your programme that you hope will make the changes sustainable?
Session 4: Discussion of EQ 2 **30 minutes**	**How significant was BCURE's contribution to priority outcomes, alongside the contribution of non-** **BCURE factors?** *a) What is the evidence that BCURE contributed to causing the observed change, and what is the evidence that non-BCURE factors contributed?* *b) What is the relative importance of BCURE and non-BCURE factors in explaining the observed change?*	Focus on the longer-term, priority outcomes. Ask: • How significant do you think BCURE's contribution was to this? • What else was going on, that might have contributed to this outcome? Use insights from the document review and PEA analysis as prompts (e.g. other programmes, supportive elements within the context, etc.)
Session 5: Discussion of EQ 3	**How and why did BCURE contribute or fail to contribute to priority outcomes?**	In advance of the workshop, write up prompts for relevant ICMOs on flipchart / print them out on A3 [A3 prompts saved in Dropbox].

Session	Details	Instructions for facilitator
1 hour	*a) Through which mechanisms, enabled by which features of the intervention and features of the (individual, interpersonal, organisational and institutional) context, did BCURE contribute to the observed change? (Testing and confirming/rejecting CIMO configurations)* *b) Where hypothesised change did not happen, or where BCURE did not contribute to observed change, how and why was this the case (through which mechanisms, features of the intervention and features) of the context)? (Testing and confirming/rejecting CIMO configurations)*	Starting with the priority outcomes, ask: How and why do you think the programme contributed / struggled to contribute to this change? Write up on post-its / flipchart, or annotate our CMO prompts to show how intervention leads to change • *What is it about the programme that led to the change?* • *What did the programme provide that was new? (Information, skills practice, opportunities for collaboration, technical support, access to evidence sources, etc.?)* • *What is it about the way the programme is implemented that made a difference, or failed to?* • *What is it about this place / context that makes the intervention work or made it not work?* • *Has change happened in the same way for all participants? In what ways it differed, and for whom? What is driving these differences?* *Bring CIMO prompts into the conversation when relevant, and ask the team to reflect on them. Aim to test all of the CIMOs identified as a priority in the evidence mapping process.*
Close **10 mins**	**Wrap up and final reflections** What else do you think we need to know, to really understand how this program has worked here?	Thanks for participation Repeat what we'll do with the info

7.3 Programme participant topic guide

Case study and country	
Interviewee name	
Position and organisation	
Interviewer name	
Date of interview	

Introduction

- We are independent researchers investigating the [xxx] project, which is funded by the UK Department for International Development. We want to hear your thoughts on this project.
- The interview will last about 1 hour.

Consent

- Everything you tell us will be confidential, and your name will not be used in any of our reports. However, we would like to use your thoughts and some anonymised quotes from the interview in our findings, if you are happy with this?
- Do you mind if we audio record the interview? This is for the researchers' reference and will allow us to check that we have we recorded your views correctly.
- Do you have any questions about the research, or concerns you would like to raise before we start?

Aim of the interview

- We'd like to talk to you about what has changed, if anything, since the start of the programme. However, this interview might be slightly different to others you may have had in the past.
- We're not just interested in whether the programme has been successful – we want to know how and why. So I'm very interested in your ideas about how and why things have changed, or not.
- We have some initial ideas but we're not sure if they are correct or not, so we will share these with you during the interview and get your thoughts.

Role and involvement in policy / decision making

- Could you please introduce yourself and your role within the organisation?
- Can I briefly check – how would you describe your role in relation to [or how are you involved in] policy and decision making? Can you summarise that for me please?
- Can you tell me what your involvement in (or contact with) this programme has been?
    - *What were the specific activities and when did you participate?*
    - *When was your first contact, and when was your last contact with the programme?*

Description of project (for stakeholders without much knowledge of it)

- The [xxx] programme aims to [encourage the use of evidence in policy and decision making / insert local description here].
- In [xxx country] [insert local description here: e.g. AFIDEP has been leading the SECURE Health programme, providing training and coaching to staff in parliament and MoH, providing technical support to health policy, convening science policy cafes and strengthening policy and research networks.]
- The programme as a whole is funded by DFID and worked across 11 countries in Africa and Asia.
- We are evaluating the programme in order to investigate how effective the programme has been, but also to understand more about how and why different types of approaches can help support evidence-informed policymaking in different contexts.

EQ 1 and 2: Outcomes and contribution

- What do you consider the outcomes of the programme to have been for you personally?
- What do you consider the outcomes of the programme to have been for [the organisation / Ministry]?

  - *Probes: last year you said that xxx had happened. Is this still the case?*
- What do you think caused these changes?
- Apart from the BCURE programme, has anything else contributed?
  - *Probe for other factors using other PE prompts, e.g. other initiatives; we have heard that the President is keen on EIPM, to what extent so you think this has been a factor*

Insert outcome probes here. Use them to make sure all of the priority outcomes for this respondent have been explored.

- Are there things that **did not** happen as a result of the programme [that you hoped would], or results that were **more limited** than you hoped?
- Do you think that the outcomes have been the same for all [people within the specific stakeholder group – e.g. trainees, mentees]? In what ways have they been different?
  - *Probe for examples of people who have been less engaged with the programme / haven't felt the benefits of the programme. Attempt to get names.*
- Have you noticed any differences in outcomes that relate to gender?

Policies being directly influenced by evidence

- [Where respondents have given examples of changes to practice] Do you know of any examples where a policy or bill (that you've worked on since you've taken part in the training?) has been directly or indirectly influenced by evidence?
- Can you tell me a little about it and who was involved please? *[Ask some of the following questions if time, prioritising questions that can only be answered by this respondent.]*
  - What was the purpose / goal of the policy / bill?
  - What was the outcome?
  - Did the training / support from ZeipNET feed in? How?
  - What were the other drivers of success?
  - Did it face any obstacles or blockages?
  - Who else was involved? (Government stakeholders, civil society?) – can you give us names / contact details?
  - Are there any documents we could look at?

EQ 3: How and why did BCURE contribute or not contribute?

For each of the changes and non-changes mentioned throughout the interview:

- You said that xxx has happened and that the programme contributed to that. Why do you think the programme made a difference to xxx?

  OR
- You said that xxx hasn't changed / the programme didn't contribute. Why do you think the programme **didn't** influence this?

Insert CIMO tables here. Use them to make sure all the CIMOs prioritised for this respondent have been covered

Longer-term outcomes/sustainability

You may not need to ask these questions separately. It may be possible to ask about 17 a, c and d in one question, as they are interlinked. Always ask about outcome 17a as this applies in all BCURE contexts. Ask about the other outcomes where relevant activities have been undertaken / results observed within the programme.

Outcome 17a	You've mentioned x, y, z [in relation to individual behaviour change]. How consistently do you think officials [in xxx department] use evidence [in their day-to-day work / when developing [policy briefs, concept notes etc.]? How far would you

Always ask this question	say the [department] is along this journey from evidence use being ad hoc vs being more consistent and routine? What else needs to happen for this to become ***sustained*** as a routine? What other factors might influence this? E.g. role of external actors on policy analysis and scrutiny (e.g. international donors; lobby groups; civil society groups)?
Outcome 17b *Ask if there have been activities and results relating to consultation process etc.*	You've mentioned x, y, z [in relation to policy development processes.] How far do you think policy processes in [the parliament] in general engage with evidence from different stakeholders and perspectives? What else needs to happen for this to become ***sustained*** as a routine? E.g. What are the mechanisms for consultation, participation and inclusion in policy processes and the way in which citizens are involved in policy development and monitoring (e.g. referendum, opinion surveys)?
Outcome 17c *Ask if there have been activities and results relating to evidence tools*	You've mentioned x, y, z [in relation to tools]. Are these tools being consistently used by people? Or is the use of the tools more ad hoc? What else needs to happen for this to become ***sustained*** as a routine? What other factors might influence this?
Outcome 17d *Ask if there have been activities and results relating to evidence processes / standards*	You've mentioned x, y, z [in relation to processes / standards]. Do you think they are helping to promote consistent evidence use by people in [the Ministry / x department]? Are they supported by senior managers? What else needs to happen for this to become ***sustained*** as a routine? What other factors might influence this?

Wrap up

- Finally, if you could change something about this [intervention] to make it work more effectively here, what would you change and why?
- Is there anything else you think we should know about the programme that we haven't already covered?
- [If you need to identify additional stakeholders for interviews] We are interested in speaking to a number of people from [xxx department], but also some people from outside the department, to give us a 360 degree picture of how the programme has interacted with it.
    - Can I please check with you which of your colleagues also participated in the programme?
    - Who is your line manager / who manages the [xxx unit]
    - Which other units / departments do you work with regularly? Can you suggest anyone in these units / departments I could speak to?

Thank respondent for their time, remind them about any documents they said they could share with you, and ask them if they would mind you getting in touch again if you have any follow-up questions.

Interviewers' reflections on interview *(consider respondent's attitude towards interview / programme; potential issues that may affect how much weight to give claims made by respondent such as motivations, plausibility of claims, inconsistencies in respondent's account):*

7.4 Programme non-participant topic guide

Case study and country	
Interviewee name	
Position and organisation	
Interviewer name	
Date of interview	

Introduction

- We are independent researchers investigating the [xxx] project, which is funded by the UK Department for International Development. We want to hear your thoughts on this project.
- The interview will last about 1 hour.

Consent

- Everything you tell us will be confidential, and your name will not be used in any of our reports. However, we would like to use your thoughts and some anonymised quotes from the interview in our findings, if you are happy with this?
- Do you mind if we audio record the interview? This is for the researchers' reference and will allow us to check that we have we recorded your views correctly.
- Do you have any questions about the research, or concerns you would like to raise before we start?

Aim of the interview

- We'd like to talk to you about the role of evidence in policymaking in [xxx sector] AND / OR [specific outcomes the respondent should have an insight into]. However, this interview might be slightly different to others you may have had in the past. I'm not just interested in what is happening, but also in your ideas about how and why things have changed, or not, over the past few years
- We have some initial ideas but we're not sure if they are correct or not, so we will share these with you during the interview and get your thoughts.

Role and involvement in policy / decision making

- Could you please introduce yourself and your role within the organisation?
- Can I briefly check – how would you describe your role in relation to [or how are you involved in] policy and decision making? Can you summarise that for me please?
- Have you heard about the [xxx] programme?

Description of project (for stakeholders without much knowledge of it)

- The [xxx] programme aims to [encourage the use of evidence in policy and decision making / insert local description here].
- In [xxx country] [insert local description here: e.g. AFIDEP has been leading the SECURE Health programme, providing training and coaching to staff in parliament and MoH, providing technical support to health policy, convening science policy cafes and strengthening policy and research networks.]
- The programme as a whole is funded by DFID and worked across 11 countries in Africa and Asia.
- We are evaluating the programme in order to investigate how effective the programme has been, but also to understand more about how and why different types of approaches can help support evidence-informed policymaking in different contexts.

Political economy analysis discussion [could have this conversation up-front, towards the end, or interspersed throughout the interview]

- I'd like to get your thoughts on the goal of [programme's] work. xxx is trying to promote better use of evidence in policymaking, through building the capacity of civil servants. What are your thoughts on this goal in the [country context]?
- Have you worked or come into contact with the [specific sectors of interest in the case study]? Other than capacity, do you have any insights into the main dynamics that affect evidence use in the [xxx sectors]

Insert questions from the PEA template here. Make sure all of the key questions have been covered, prioritising questions that are difficult to answer through secondary document review.

- Who else do you think we could speak to, to get an insight into the political dynamics in xxx sector? Do you have any documents or sources that we could draw on?

EQ 1: Outcomes and sustainability

- [If respondent has heard about the programme] What is your general impression of the xxx programme?
- Over the past three years, have you noticed any changes in the way [xxx sector / department] thinks about or uses evidence in decision making? What kinds of changes? Can you give me an example?

Insert outcome probes here. Use them to make sure all of the priority outcomes / ToC links prioritised for this respondent have been explored.

- Do you think that the changes have been the same for all [people within the specific stakeholder group – e.g. senior stakeholders]? In what ways have they been different?
    - *Probe for examples of people who have been less engaged / have resisted change. Attempt to get names.*
- Are there things that **are not changing** in relation to how [xxx sector / department] uses evidence? Or changes that are happening more slowly?

EQ 2: What was BCURE's contribution to observed changes?

For each of the changes mentioned under EQ 1:

- You said that xxx has happened / changed. What do you think caused that change?
- Apart from the BCURE programme, has anything else fed into this?
    - *Link back to the initial PEA discussion*

EQ 3: How and why did BCURE contribute or not contribute?

For each of the changes and non-changes mentioned under EQ 1:

- You said that xxx has happened and that the programme contributed to that. Why do you think the programme made a difference to xxx? **OR** you said that xxx hasn't changed / the programme didn't contribute. Why do you think the programme didn't influence this?

Insert CIMO tables here. Use them to make sure all the CIMOs prioritised for this respondent have been covered

Policies being directly influenced by evidence

- Have any of these changes you've mentioned led to a specific policy or bill being influenced by evidence?
- Can you tell me a little about this and who was involved please?

Longer-term outcomes/sustainability

You may not need to ask these questions separately or at the end of the discussion. It may be possible to ask about 17 a, c and d in one question, as they are interlinked, or to ask about them under EQ 1 if the opportunity arises. Always ask about Outcome 17a as this applies in all BCURE contexts. Ask about the other outcomes where relevant activities have been undertaken / results observed within the programme

Outcome 17a	You've mentioned x, y, z [in relation to individual behaviour change]. How consistently do you think officials [in xxx department] use evidence when developing

	[policy briefs, concept notes etc.]? How far would you say the [department] is along this journey from evidence use being ad hoc vs being more consistent and routine? What else needs to happen for this to become ***sustained*** as a routine? What other factors might influence this? E.g. role of external actors on policy analysis and scrutiny (e.g. international donors; lobby groups; civil society groups)?
Outcome 17b *Ask if there have been activities and results relating to consultation process etc.*	You've mentioned x, y, z [in relation to policy development processes.] How far do you think policy processes in [the parliament] in general engage with evidence from different stakeholders and perspectives? What else needs to happen for this to become ***sustained*** as a routine? E.g. What are the mechanisms for consultation, participation and inclusion in policy processes and the way in which citizens are involved in policy development and monitoring (e.g. referendum, opinion surveys)?
Outcome 17c *Ask if there have been activities and results relating to evidence tools*	You've mentioned x, y, z [in relation to tools]. Are these tools being consistently used by people? Or is the use of the tools more ad hoc? What else needs to happen for this to become ***sustained*** as a routine? What other factors might influence this?
Outcome 17d *Ask if there have been activities and results relating to evidence processes / standards*	You've mentioned x, y, z [in relation to processes / standards]. Do you think they are helping to promote consistent evidence use by people in [the Ministry / x department]? Are they supported by senior managers? What else needs to happen for this to become ***sustained*** as a routine? What other factors might influence this?

Wrap up

- Is anything else you think we should know about the programme that we haven't already covered?
- If there anything else you think we need to know, to really understand the role that evidence plays in policymaking in this context?

Thank respondent for their time, remind them about any documents they said they could share with you, and ask them if they would mind you getting in touch again if you have any follow-up questions.

Interviewers' reflections on interview *(consider respondent's attitude towards interview / programme; potential issues that may affect how much weight to give claims made by respondent such as motivations, plausibility of claims, inconsistencies in respondent's account):*

7.5 Programme staff topic guide

Case study and country	
Interviewee name	
Position and organisation	
Interviewer name	
Date of interview	

Introduction

- We'd like to talk to you about your perceptions of the BCURE programme. As you know, we're not just interested in what is happening, but also in your ideas about how and why things have changed, or not, over the past few years. We'd like to share our initial ideas with you during the interview and get your thoughts.
- The interview will last about 1 hour.

Consent

- Everything you tell us will be confidential, and your name will not be used in any of our reports. However, we would like to use your thoughts and some anonymised quotes from the interview in our findings, if you are happy with this?
- Do you mind if we audio record the interview? This is for the researchers' reference and will allow us to check that we have we recorded your views correctly.
- Do you have any questions about the research, or concerns you would like to raise before we start?

Role and involvement in BCURE

- Could you please introduce yourself and your role within the organisation?

Political economy analysis discussion

We are interested in understanding the national and sector-level political dynamics that affect policymaking in [xxx] sector.

- In your opinion, what are the main issues and dynamics that affect evidence use in [xxx sector / Ministry]?

Insert questions from the PEA template here. Make sure all of the key questions have been covered, prioritising questions that are difficult to answer through secondary document review.

- Who else do you think we could speak to, to get an insight into the political dynamics in xxx sector? Do you have any documents or sources that we could draw on?

EQ 1: Outcomes and sustainability

- What do you consider the outcomes of the programme to have been for [each of the main stakeholder groups the programme is working with]? Can you give examples?

Insert relevant outcome probes here, attempting to build on the insights from the programme team workshop / plug gaps.

- Do you think that the changes have been the same for all [people within the specific stakeholder group – e.g. senior stakeholders]? In what ways have they been different?
    - *Probe for examples of people who have been less engaged / have resisted change. Attempt to get names.*
- Are there things that **did not** happen as a result of the programme [that you hoped would], or results that were **more limited** than you hoped?

EQ 2: What was BCURE's contribution to observed changes?

For each of the changes mentioned under EQ 1:

- What role do you think BCURE played in promoting [xxx change]?
- Apart from the BCURE programme, has anything else fed into this?
    - *Link back to the initial PEA discussion*

EQ 3: How and why did BCURE contribute or not contribute?

For each of the changes and non-changes mentioned under EQ 1:

- You said that xxx has happened and that the programme contributed to that. Why do you think the programme made a difference to xxx? **OR** you said that xxx hasn't changed / the programme didn't contribute. Why do you think the programme didn't influence this?

Insert CIMO tables here, attempting to build on the insights from the programme team workshop / plug gaps

Policies being directly influenced by evidence

- Have any of these changes you've mentioned led to a specific policy or bill being influenced by evidence?
- Can you tell me a little about this and who was involved please?

Longer-term outcomes/sustainability

You may not need to ask these questions separately or at the end of the discussion. It may be possible to ask about 17 a, c and d in one question, as they are interlinked, or to ask about them under EQ 1 if the opportunity arises. Always ask about Outcome 17a as this applies in all BCURE contexts. Ask about the other outcomes where relevant activities have been undertaken / results observed within the programme

Outcome 17a	You've mentioned x, y, z [in relation to individual behaviour change]. How consistently do you think officials [in xxx department] use evidence when developing [policy briefs, concept notes etc.]? How far would you say the [department] are along this journey from evidence use being ad hoc vs being more consistent and routine? What else needs to happen for this to become ***sustained*** as a routine? What other factors might influence this? E.g. role of external actors on policy analysis and scrutiny (e.g. international donors; lobby groups; civil society groups)?
Outcome 17b *Ask if there have been activities and results relating to consultation process etc.*	You've mentioned x, y, z [in relation to policy development processes.] How far do you think policy processes in [the parliament] in general engage with evidence from different stakeholders and perspectives? What else needs to happen for this to become ***sustained*** as a routine? E.g. What are the mechanisms for consultation, participation and inclusion in policy processes and the way in which citizens are involved in policy development and monitoring (e.g. referendum, opinion surveys)?
Outcome 17c *Ask if there have been activities and results relating to evidence tools*	You've mentioned x, y, z [in relation to tools]. Are these tools being consistently used by people? Or is the use of the tools more ad hoc? What else needs to happen for this to become ***sustained*** as a routine? What other factors might influence this?
Outcome 17d *Ask if there have been activities and results*	You've mentioned x, y, z [in relation to processes / standards]. Do you think they are helping to promote consistent evidence use by people in [the Ministry / x department]? Are they supported by senior managers?

relating to evidence processes / standards	What else needs to happen for this to become ***sustained*** as a routine? What other factors might influence this?

Wrap up

- Is anything else you think we should know about the programme that we haven't already covered?
- If there anything else you think we need to know, to really understand the role that evidence plays in policymaking in this context?

Thank respondent for their time, remind them about any documents they said they could share with you, and ask them if they would mind you getting in touch again if you have any follow-up questions.

Interviewers' reflections on interview *(consider respondent's attitude towards interview / programme; potential issues that may affect how much weight to give claims made by respondent such as motivations, plausibility of claims, inconsistencies in respondent's account):*

8 Analysis frameworks

As described in Section 3, analysis frameworks were developed in Excel to guide a systematic and transparent analysis of the evidence against the EQs, at country case study and synthesis level.

Country case study analysis database template

	Background details					EQ 1: what happened, for whom?				EQ 2: BCURE contribution			EQ 3: How and why (CIMOs)						Other
#	Interview respondent/ document name	Reflections on potential bias / position in relation to programme	Stakeholder type	Relevant intervention	Organisation	Outcome no.	Outcome	Evidence	Reflection on strength of evidence	Detail of BCURE contribution	Details of other factors	Reflection on strength of evidence / bias	EQ 3 hypothesis (CIMO)	Explicitly tested or inferred?	CIMO confirmed, rejected, revised, or new CIMO suggested?	Evidence for intervention factors	Evidence for contextual factors	Mechanism	Any other thoughts or comments

Overview report synthesis database template

Background details			EQ 1: what happened, for whom?				EQ 2: BCURE contribution			EQ 3 - How and why did BCURE contribute? (CIMOs)							Other	
Case	Setting / impact pathway	Level *individual, organisational, institutional?*	Outcome	Extent outcome achieved	Evidence for outcome (narrative and quotes)	Reflection on strength of evidence	EQ 2 - BCURE contribution - details (narrative and quotes)	Details of other factors, factors)	Reflection on strength of evidence / bias	EQ 3 hypothesis (CIMO)	Evidence for intervention factors	Evidence for contextual factors	Mechanism	CIMO confirmed, rejected, revised, or new CIMO suggested?	Revised CIMO (reflecting what actually happened at Stage 3)	Reflection on strength of evidence	Comments	Lessons

9 Additional analysis tables

Detailed summary of results from BCURE training approaches

	Training approach and scale	Improvements in individuals' capacity *Kirkpatrick Level 1*	Changes in individuals' behaviour *Kirkpatrick Level 2*
Pakistan *Evidence from Stage 3*	Large scale EIPM training course (1780 participants), limited targeting, no follow-up	Strong evidence through pre- and post-training tests of significant gains in technical knowledge. Participants and trainers felt that training quality was enhanced by the use of contextually relevant case studies, and felt some of the practical tools like cost-benefit analysis were particularly useful, although some participants at Stage 3 complained that the sessions were quite academic and not particularly interactive	Limited evidence of widespread behaviour change, although the evaluation was only able to interview a small proportion of the whole cohort. Where trainees had applied their learning, the training had generally helped trainees frame their thinking on how to use evidence or data to address a specific task in the workplace (M), and gave them tools (e.g. cost-benefit analysis) to facilitate specific types of analysis (M) in a context where they were tasked with working on an issue that required data (C), where senior managements were supportive and actively encouraged them (C), and where they were able to draw on resources to access and analyse evidence in response to a specific problem (C). This was enabled by participants' existing skills, motivation and experience, including soft skills that enabled them to present analysis to senior members of the government to influence reform (C). Trainees who hadn't applied their learning often did not perceive the training as relevant to their current professional role, often because they were not involved in policy formulation (C). This was a consequence of the training being delivered to whole cadres of staff as part of mandatory requirements for promotion rather than based on a decision about relevance and need (I). Interviews also suggested missing incentives in the workplace to change practices towards more evidence-informed policymaking, including uninterested senior managers themselves lacking an incentive to consider evidence (C), against a backdrop of endemic corruption providing a motive to ignore or suppress evidence (C)
Bangladesh *Evidence from Stage 3*	Large scale EIPM training course (400 participants), for civil servants involved in policy formulation in three pilot ministries, with content aligned to EIPM guidelines developed at Cabinet Division Level. On-the-job follow-up support to a small number of trainees, to apply guidelines in policy pilots	Strong evidence of increased knowledge and understanding of EIPM, through pre- and post-training tests corroborated by interviews Participants felt the training was useful and high quality because it provided a systematic procedure for policymaking, it was practical and incorporated local case studies, and the trainers were local experts with relevant experience	Strong evidence that training led to new or improved knowledge, skills and confidence and improved trainees' self-efficacy (M), leading to changes in the way evidence was considered in policy formulation (O), where trainees were supported to apply skills through policy pilots after the training (I). A key factor was the fact that the EIPM guidelines were seen as providing a helpful structure to facilitate policy formulation (M) in a context where such guidance was lacking (C). Limited evidence (as yet) that trainees who did not receive follow-up support had had an opportunity to apply skills (C) – although the top down EIPM guidelines are likely to facilitate skills application (M) if they are adopted by line ministries and trainees are requested to use them (M). However, some trainees were not involved in policy formulation roles, suggesting there may be limited opportunities to apply learning in future (C). Several stakeholders suggested that without follow-up (for example refresher training, or permanent EIPM focal points within ministries) (I) there is a risk that trainees will forget what they learned or will lack the confidence to apply their learning (O)
Zimbabwe	Small scale EIPM training course (49 participants) delivered to	Strong evidence (though training follow-up survey and interviews with	Strong evidence for sustained change in the work of trainees in the Ministry of Youth as a result of BCURE. Training generated self-efficacy (M) and contributed to sustained behaviour change (O) because it helped trainees perform in their new

	Training approach and scale	Improvements in individuals' capacity *Kirkpatrick Level 1*	Changes in individuals' behaviour *Kirkpatrick Level 2*
Evidence from Stage 3	technical staff in targeted institutions, followed up by technical support to implement organisational reforms and a Parliamentary exchange programme	significant number of participants) that trainees had gained new knowledge and skills (although less so for participants outside the research department in parliament). Participants in the Ministry of Youth felt the training was high quality because it was practical, hands on and participatory, and because it imparted soft skills as well as technical skills	roles as research officers in a newly-established research unit (C), as most did not have background in research (C), and given the resource-constrained context where other training was unavailable (C). Skills application was supported by the two Directors in the unit, who were interested in evidence and supportive of trainees applying their skills (C). However, opportunities to apply skills in research work were limited by the small scale of the research unit – which is shrinking due to staff rationalisation (C), and the fact that it interacts with only part of the Ministry (C), and that officers are often engaged in administrative rather than research work (C). Limited evidence that the training or exchange programmes made a significant contribution to changes in practice in parliament (O). The training content was insufficiently tailored to the parliament-specific needs of staff (I) (particularly those outside the research department), and therefore was not relevant to trainees' needs (C). The plethora of other training and exchange opportunities available in parliament may also explain why some participants felt the training or exchange visits did not offer much that was new (C)
Kenya *Evidence from Stage 3*	Small scale EIPM training course (45 participants), delivered to technical staff in targeted institutions, followed up by mentoring support and an overseas secondment	Strong evidence (through training follow-up survey and interviews with significant number of participants) of increases in knowledge and skills. Participants felt the training was useful and high quality because it combined theory with practical application and provided the space to work on a live policy topic, and the facilitators were high calibre, knowledgeable, patient, skilled and committed; although some felt that course had been insufficiently tailored for Parliamentary staff	Strong evidence (from triangulated interviews with participants and managers) that substantial numbers of trainees in both parliament and the MoH had been able to use learning in their work, and this behaviour change had been sustained up until the final evaluation. In parliament, training succeeded in building self-efficacy (M) which resulted in improved use of evidence (O) in a context where training was delivered to a newly-recruited researchers during their induction periods (I), helping them quickly meet the specific demands of their jobs (C). Skills application was supported by follow-up support from BCURE to produce concrete evidence products, tools and templates (I), which facilitated (M) trainees to more efficiently meet the high volume of evidence products required of them (O), reinforced by senior managers in the unit (M), who were already proactively engaged in an evidence agenda (C), by providing hands on support and feedback and encouraged trainees to improve the quality of their work (C), which in turn generates recognition and career rewards (C) that increase motivation for evidence use (M).' In the Ministry of Health, training had also succeeded in building self-efficacy (M), and the searching, synthesising and presentation approaches and tools provided by BCURE helped facilitate (M) trainees to present evidence more effectively. These contributed to improved evidence use (O) particularly among trainees with opportunities to apply evidence in the development of specific policies, standards and guidelines (C). The most sustained gains in individual capacities and ongoing use of evidence (O) was among motivated individuals (C) based in divisions and units with well-resourced donor programmes that offer trainees opportunities to apply and further develop EIPM skills, which in turn generates recognition and career rewards (C) that increase motivation for evidence use (M). Evidence use was less sustained in divisions where officials take a more administrative role in policy development and there are fewer opportunities to apply skills (C)
Sierra Leone	Medium scale training (964 training days),	Strong evidence through Stage 2 interviews that	Some evidence at Stage 2 (from interviews with participants and programme staff) that line ministry Cabinet Focal Persons trained to support line ministries to apply new cabinet

	Training approach and scale	Improvements in individuals' capacity *Kirkpatrick Level 1*	Changes in individuals' behaviour *Kirkpatrick Level 2*
Evidence from Stage 2	delivered to technical staff in line ministries to help support implementation of organisational reforms to cabinet procedures. Training also delivered to cabinet staff to help them support implementation	training and regular meetings had helped civil servants understand the new procedures established in the new cabinet manual. Participants appreciated the participatory nature of the training, the use of practical case studies and the opportunity to learn and share from colleagues	Procedures were able to perform their new roles to some extent, but that there was a 'long way to go.' Training, combined with ongoing support from the new (BCURE supported) cabinet research unit (I) helped increased cabinet focal person self-efficacy (M) to perform the duties of their new role and support the implementation of the procedures (O). The presence of the Cabinet Secretary in the training helped ensure full participation, especially of senior civil servant staff (I). However, the need for support from other ministry staff for CFPs to perform their role (C), potentially undermines their ability to apply their learning
South Africa *Evidence from Stage 2*	Small scale workshops, aiming to provide an introduction to EIPM to participants who might become mentees, raise awareness about EIPM and its potential value, and provide spaces for dialogue	Some evidence through Stage 2 interviews that workshops had introduced participants to relevant terminology and methods and reinforced their understanding of the importance of evidence. Participants appreciated the opportunities to share challenges and solutions in the workshops, although there was limited time for practical skills application	Limited evidence of instrumental changes in practice as a result of the workshops. Where there was strong prior interest and enthusiasm for EIPM (C), and where it was relevant to existing work (C), this seemed to enable workshop attendees to connect to the concepts and understand the immediate usefulness to their work (M), assisting in conceptualising their work and offering potential solutions to work challenges (O) Where prior interest and understanding were absent (C), the workshops increased interest and awareness in EIPM and the workshop content was regarded as potentially useful (M), but participants were not actively applying the concepts and methods (O)

10 Communications framework

This section summarises the evaluation communications framework, developed during the inception phase and updated during the mid-point of the evaluation.

Summary
The primary aim of the BCURE evaluation is to strengthen the global evidence base on whether capacity-building approaches to support evidence-informed policy can be a cost effective way to reduce poverty, and if so how they can be implemented to achieve the greatest impact. This framework describes how communications will support the evaluation in achieving this outcome.

The role of the BCURE evaluation communication function is to carry out activities that not only raise awareness of the evaluation with target audiences but also help best position the learning and findings for uptake. By ensuring that lessons learned from the evaluation on what works and doesn't are strategically shaped and shared, we hope that funders, designers and implementers of evidence-informed policymaking (EIPM) initiatives can make better choices when it comes to supporting similar initiatives.

The communications function is also responsible for sharing what is understood about the effectiveness of the BCURE programme with DFID and its implementing partners.

The BCURE evaluation communication strategy contains the following objectives:

1. To communicate where and how, and in what circumstances, decision makers can better access, use and understand evidence
2. To provide support and assistance to the BCURE programmes on the most effective ways of communicating evaluation findings to partners and key audiences in the countries in which they work
3. To reach, engage and inform the emerging Community of Practice around EIPM about how and why capacity-building for evidence use is important and effective in improving development outcomes.

This **communications framework** supports these objectives by identifying and analysing the evaluation's target audiences, identifying the opportunities and spaces for engagement and planning the specific activities and channels that will be used.

The latest **situation analysis** conducted as part of this framework highlights a number of difficulties in identifying both 'new' audiences and the spaces for sharing the evaluation findings. For example, although the EIPM 'community' is a key entry point for engagement, it is by and large led by 'supply-side' actors such as researchers, donors and knowledge brokers rather than important 'demand-side' actors such as the high and mid-level government policymakers who are at the heart of BCURE CToC. The analysis also finds that while there are a number of health- and development research-focused EPIM and capacity development networks and initiatives, there are relatively few that focus specifically on governance and public sector reform in developing countries– a key field for the uptake of the evaluation findings.

Since many of the EIPM platforms and spaces for engagement are largely driven by a core number of actors (including BCURE implementing partners INASP and AFIDEP), we conclude that the most effective way to reach the majority of our stakeholders is by utilising existing channels (such as knowledge platforms and networks) and through the BCURE implementing partners and DFID's Evidence into Action team who we have identified as **'amplifiers'** as well as recipients of the evaluation findings.

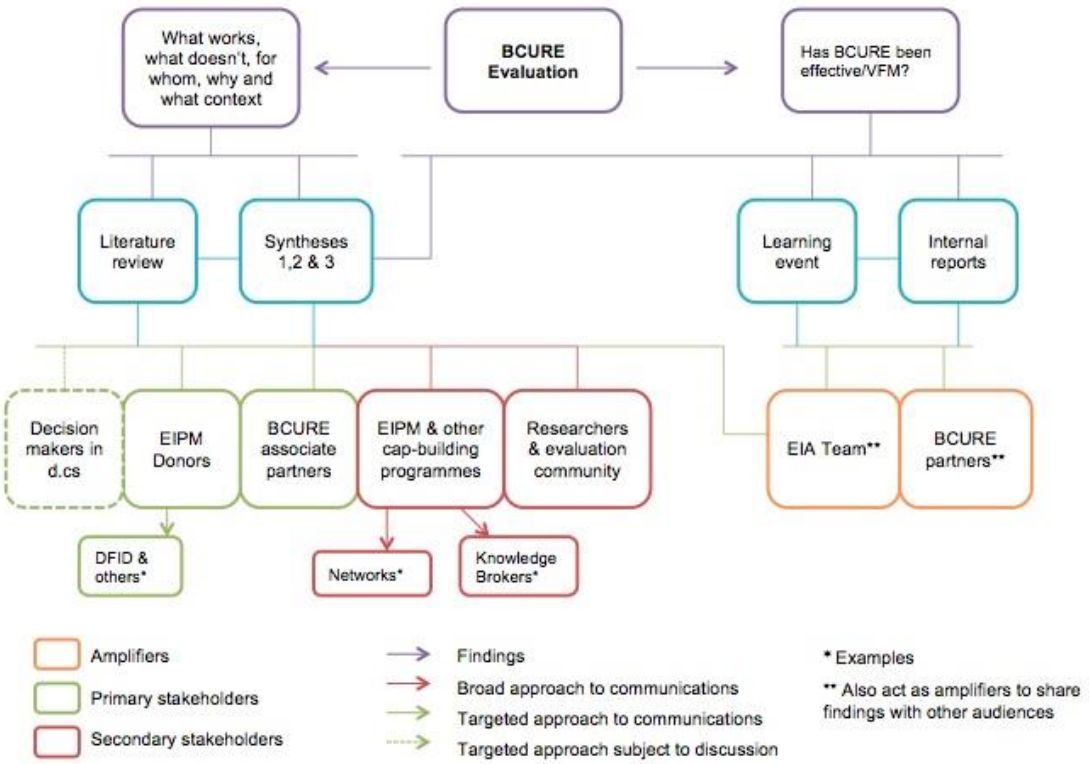
The latest round of **stakeholder identification and analysis** has begun to fill the gaps in our knowledge of key target groups, such as governance programmes, additional EIPM donors and their portfolios. In the process we have found that while the BCURE evaluation is well equipped to share top-line messages with a general set of audiences, a more nuanced understanding of our primary target stakeholders is needed so that findings can be packaged and channelled in ways that are appropriate to their needs.

To address these issues, we propose taking a **two-pronged approach** to communicate with three tiers of target audiences (Figure 1). To maximise effectiveness, we suggest implementing our outreach in close collaboration with a group of **amplifiers**. This will be a group of selected individuals who will help create demand for findings from the evaluation among primary and secondary target audiences by profiling the evaluation findings and encouraging debate on the findings.

Our most intensive engagement will be with a set of **primary target audiences** who require a more **differentiated approach**. For these, we plan to use a range of well-tailored communications products and channels to regularly communicate the evaluation findings including via our amplifier intermediaries. For **secondary audiences**, our engagement will take a much **broader approach** by utilising multiple communication channels (including intermediaries such as networks and knowledge platforms) but less frequently and with less tailoring to meet specific informational needs.

Figure 1 illustrates our **two-pronged approach** to BCURE communications that focuses on both broad and targeted approaches according to the different stakeholder groups. The following sections then discuss aspects of the communications framework in more detail.

Figure 5: BCURE evaluation communications approach



Situation analysis for BCURE evaluation communications

This area of work seeks to establish:

1. The resources and communication channels the BCURE evaluation has for capturing learning and undertaking dissemination during the programme
2. The likely sources of information and opportunities to share learning
3. Dissemination opportunities (audiences/spaces) during the life of the programme relating to the evaluation (e.g. online platforms, conferences, online discussions)

The situation analysis for this framework builds on work already done as part of the communications strategy and looks at the challenges and opportunities presented to BCURE evaluation communications specifically rather than to the programme per se.

Information in relation to this was gathered during the inception period and supplemented with further desk-based research for the purposes of this document. Sources used include BCURE and BCURE evaluation documentation (including the common theory of change); a review of external barriers and opportunities (see communications strategy), the stakeholder analysis to date (also see communications strategy; websites of organisations and networks working in EIPM; and records of discussions with members of the BCURE evaluation team and DFID. A full list of channels and opportunities is available on request.

Table 6: Challenges and opportunities for BCURE evaluation communications

Challenge	Opportunity
The concept of EIPM in target countries is relatively new and therefore relatively low profile	Use amplifiers to help create demand for findings from the evaluation among primary and secondary target audiences by profiling the evaluation findings and encouraging debate on the findings. For example, the evaluation senior adviser Professor Fanie Cloete, at Stellenbosch who has a high profile in EIDM/PM in Africa. Spot opportunities beyond these spaces for sharing evaluation findings e.g. DFID Governance Advisers retreat, What Works Global Summit etc. Keep informed of relevant debates and discussions on EIPM by plugging into and engaging with knowledge platforms and discussion fora such EPBDN, the Pelican Initiative, Policy & Ideas, Knowledge Brokers Forum and the BCURE DGroups.
There are few EIPM initiatives tackling the issue beyond health and research uptake	Broaden outreach by identifying programmes and networks in which capacity building towards EIPM is a component. Engage with stakeholders moving into the government space through research-policy networks such as the UKCDS Research Strengthening Group and the Think Tank Initiative. Encourage DFID to showcase BCURE evaluation as an example of accountability and transparency, and building effective government institutions.
Capacity development networks are often generic and too broad in terms of areas of interest	Target networks that focus on building government and public sector effectiveness, where there is potential to position EIPM as relevant to public sector reform. For example, LenCD's 'Effective Institutions' working group that aims to prepare members' governments for meeting the Busan Aid Effectiveness targets; the capacity-building section of the GSDRC website; and accountability and transparency initiatives such as the international Open Government Partnership.
There are a number of assumptions underpinning the concept of 'evidence-informed policymaking'	Use existing EIPM spaces and places to convene discussions and showcase new findings such as the Alliance for Useful Evidence, Capacity Alliance and Research 2 Action. Raise awareness of the BCURE literature review insights and learning with a wide range of audiences to stimulate interest and debate. Share internal learning through external blogs and professional networking sites. For example, blogging on LEN CD, R2A, better evaluation etc.
BCURE evaluation has to build reputation/credibility with target audiences	Develop language across all communications that positions the BCURE evaluation as a unique opportunity to gain insight into the effectiveness of EIPM capacity-building initiatives. Share information about BCURE evaluation findings through the communication channels of key partners to mutually shared stakeholder groups. Utilise entry points into key EIPM initiatives via reputation and convening power of amplifiers e.g. INASP and DFID.

Research uptake by decision makers is often regarded as a simple, linear process	Interact with the growing body of supply-side organisations and networks raising awareness of the complexity of EIPM such as UKCDS. Promote the broader view of how research findings make their way into government decision making. For example the literature review briefing
Communication activities and ambition are restricted by modest budget allocation	Use low and no-cost online communication tools to share the findings including Itad website, Twitter, Tumblr, Mailchimp etc. Look for efficient ways in which harnessing capacity of BCURE evaluation team and partners to roll out certain tasks.

Stakeholder mapping and analysis

The BCURE evaluation will only realise its intended value if the findings from the programme are effectively communicated to identified audiences who then act on the new knowledge. The aim of our work in this area has been to establish the needs, interest and contexts of the stakeholders of the communication activities (e.g. the BCURE team, BCURE implementing partners DFID, other donors and organisations with an interest in EIPM).

We have identified the following three broad audience categories. Depending to their role, some stakeholders fall under more than one category.

Amplifiers

The role of amplifiers will be to:

1. Assist in creating demand for findings from the evaluation among primary and secondary target audiences by profiling the evaluation findings and what they will offer;
2. Amplify findings from the evaluation among primary and secondary target audiences by profiling communications products and events and encouraging debate on the findings.

The role of these amplifiers is critical to the outreach of the evaluation findings for a number of reasons. Firstly, it will open up windows and spaces for the communication of findings that the evaluation team are either not aware of or will struggle to reach, so extending the reach of the communications strategy. Secondly, it would ensure that communications products are well tailored to stakeholders, maximising their utility and therefore value for money.

Activities might include:

- BCURE programme implementers convening meetings to present and discuss the findings with government partners as part of their sustainability planning.
- DFID facilitating discussions with DFID country offices where BCURE interventions happen and in which the evaluation case studies are carried out.
- BCURE programme implementers encouraging partners to sign up to the BCURE evaluation newsletter.
- DFID engaging with DFID country offices where there is no BCURE programme but where there is interest in demand-side work e.g. Nepal.
- A BCURE evaluation presentation at the BCURE learning event to explore this role in greater detail.
- BCURE programme implementers cross-posting evaluation blogs on their own websites.

Regular communication with this set of stakeholders is essential. Amplifiers will be reached via BCURE DGroups and will be kept up-to-date on the work of the evaluation via a newsletter three times a year. The literature review and synthesis briefings will act as a tool for the amplifiers to take forward their own informing and influencing. Each time an evaluation product is published, amplifiers will receive a 'What's new?' e-alert with a specific call to action depending on the nature of the product. For example, these 'calls' could include asking them to share the case studies with their networks, reflect on their own learning as part

of their project blog or to host on- or off-line discussions with their own partners that explore the implications of the findings.

Primary stakeholders

The **primary target audiences** are BCURE implementing partners, BCURE implementing associates, EIPM donors and funders and government ministries and organisations. As such, they are perceived to be the most receptive to the evaluation findings and among the most influential in terms of using these in their own policy and programming. Our primary stakeholders include stakeholders who are harder to reach but whose uptake of the findings would make a significant contribution to the theory of change.

We would like these stakeholders to respond by:

- Formally registering an expression of interest. For example, emailing with questions on specific issues
- Sending invitations to the evaluation team to come and give a detailed presentation of the learning and findings
- Extending the reach of the evaluation findings to their colleagues and networks. For example, by incorporating the briefings into their EIPM resources. E.g The YakoViko Evidence-Informed Policymaking Toolkit and the LenCD learning package.

These stakeholders require a deeper level of understanding and engagement through a more targeted approach via individuals i.e. our selected amplifiers and specific programmes or portfolios of work. A mix of written products (e.g. the literature review and synthesis briefings) and digital communications (e.g. the evaluation newsletter, blogs and Twitter) will be used to communicate and position the evaluations findings for uptake. We will ask amplifiers to share these and facilitate virtual and face-to-face discussions to explore them further. For example, the DFID BCURE team might hold a discussion with country offices via Yammer or host sessions at Advisers' Professional Development Conferences.

Secondary stakeholders

Secondary target audiences are wide-ranging and easy to reach en masse via knowledge brokers and platforms. They include:

- EIPM project implementers and networks: these are a diverse set of organisations but with a common focus. Priority will be given to EIPM programmes that promote governance and accountability as well as health and research uptake. As far as possible, we will use existing EIPM fora such as WHO's EVIPNET, Health Information for All, the Alliance for Useful Evidence and capacity development networks such as LEN CD and Capacity Alliance Feeds for the communication of the evaluation findings.
- Other development partners who actively support EIPM more broadly: these include governance programmes in which EIPM is a component including health promotion, environmental governance, voice and accountability and governance reform. This is a disparate group of stakeholders. To reach them, the evaluation team will rely heavily on the amplifiers who engage directly with specific organisations.
- Research organisations, programmes and think tanks involved in consortia responsible for delivering research on EIPM and capacity development. While their influence on policy and programming is important, this influence tends to be indirect and makes itself felt over a medium-term timeframe.
- Evaluation community: this includes evaluation focused organisations and evaluation specialists. Evaluation focused organisations working with DFID will be targeted through our social media work. More direct engagement with the evaluation community will be done through focal points within the evaluation community such as the European Evaluation Society and What Works Global Summit.

We expect the responses of secondary stakeholders to be 'light touch' and could include:

- Signing up to the BCURE evaluation newsletter
- Following the BCURE evaluation on social media and sharing posts/tweets with followers
- Citing the evaluation findings in literature reviews and policy briefings

Our outreach to secondary stakeholders target will focus on communicating the breadth of the findings emerging from the evaluation and encouraging discussion about these. Activities and channels will include conducting a social media campaign via Twitter, posting regular blogs, and using other knowledge platforms such as the LenCD library and the Africa Evidence Network database to raise awareness about the evaluation and draw target audiences to the evaluations' portal, hosted on Itad's website. The evaluation methodology and results will be presented in at least three public discussions including the European Evaluation Society Annual Conference, the UKCDS Research Capacity Strengthening Group and at the Centre for Development Impact.

Target audiences

Target audience Categories	Specific stakeholders
Amplifiers	DFID Evidence into Action Team BCURE implementing partners BCURE evaluation team
Primary audience	BCURE implementing partners BCURE implementing partner associates EIPM donors and funders such as • Multilateral organisations: World Bank, NEPAD, UNDP and WHO • Bilateral organisations: SIDA, IDRC, USAID, DFID, DSIG Netherlands • Philanthropics: Hewlett Foundation, Wellcome Foundation, Gates Foundation and Open Society Institute Government ministries and organisations such as the Performance Monitoring and Evaluation Department of the South African Government. Participants (interview respondents) in the BCURE evaluation
Secondary audience	EIPM project implementers, such as the Knowledge Sector Initiative (Indonesia), Supporting the Use of Research Evidence (SURE) (WHO Worldwide), NEPAD Capacity Development Programme and DFID Nepal's Evidence for Development and NICE International. Capacity development networks such as LEN CD, Capacity4Dev.eu, APDEV, Africa Cabinet Government Network and ACBF. Research capacity strengthening organisations and networks such as UKCDS, EBPDM, iDSI and the Alliance for Useful Evidence. Other development programmes that focus on EIPM for example, the Public Sector Accountability and Governance programme, ESPINN and PATHS2 in Nigeria, ESP Nepal, FLEGT and ACT in Tanzania. Research organisations, programmes and think tanks including 3ie, the Centre for Development Impact (IDS), Overseas Development Institute, Institute for Government, the Alliance Health Policy Systems Research and the Centre for Evidence and Social Innovation. Evaluation community e.g. UK, European, African and Asian Evaluation Society Members

Summary of BCURE evaluation communications activities

- Development of a knowledge page on the Itad website, containing evaluation outputs http://www.itad.com/knowledge-and-resources/bcure/
- Publication of the literature review and Stage 1 and 2 synthesis reports along with briefing notes and blogs on the Itad website, and dissemination to key audiences listed above.
- Academic publication: Punton, M., Vogel, I. and Lloyd, R. (2016b). Reflections from a Realist Evaluation in Progress: Scaling Ladders and Stitching Theory. CDI Practice Paper, 18.

- Presentations at the European Evaluation Society Conference 2016, the What Works Conference 2016, the American Evaluation Conference 2017, and the UK Realist Evaluation Conference 2015 and 2016.
- Presentations and practical sessions with implementing partners at the BCURE Learning Events 2014, 2015 and 2016.

Following the completion of the final evaluation report, the following activities are planned:

- Full design of synthesis report to maximise readability
- Blogs on the Itad website, where possible cross-posted to reach further audiences in Table 2
- Face-to-face presentations with DFID staff
- Targeted dissemination of synthesis report with primary and secondary stakeholders in Table 2
- Follow-up calls and webinar for BCURE implementing partners
- Presentation of findings at key conferences in 2018

11 RAMESES standards for realist evaluation

In 2016, a set of reporting standards were developed for realist evaluations as part of the RAMESES II Project.[13] These standards aim to improve consistency, rigour and usability of realist evaluations. The table below sets out the standards, and indicates the relevant section of the BCURE evaluation report where each standard is addressed.

No.	Standard	Relevant section of report or annexes
1.	In the title, identify the document as a realist evaluation	See title
Summary / abstract		
2.	Journal articles will usually require an abstract, while reports and other forms of publication will usually benefit from a short summary. The abstract or summary should include brief details on: the policy, programme or initiative under evaluation; programme setting; purpose of the evaluation; evaluation question(s) and/or objective(s); evaluation strategy; data collection, documentation and analysis methods; key findings and conclusions. Where journals require it and the nature of the study is appropriate, brief details of respondents to the evaluation and recruitment and sampling processes may also be included. Sufficient detail should be provided to identify that a realist approach was used and that realist programme theory was developed and/or refined.	See Executive Summary
Introduction		
3. Rationale for evaluation	Explain the purpose of the evaluation and the implications for its focus and design	See Section 1 of the main report
4. Programme theory	Describe the initial programme theory (or theories) that underpin the programme, policy or initiative	Annex 4 details the programme theory and how it has evolved over time
5. Evaluation questions, objectives and focus	State the evaluation question(s) and specify the objectives for the evaluation. Describe whether and how the programme theory was used to define the scope and focus of the evaluation	See Annex 3.1
6. Ethical approval	State whether the realist evaluation required and has gained ethical approval from the relevant authorities, providing details as appropriate. If ethical approval was deemed unnecessary, explain why	See Annex 3.10
Methods		
7. Rationale for using realist evaluation	Explain why a realist evaluation approach was chosen and (if relevant) adapted	See Annex 3.2
8. Environment surrounding the evaluation	Describe the environment in which the evaluation took place	See Section 3.1 of the main report
9. Describe programme policy, initiative or	Provide relevant details on the programme, policy or initiative evaluated	See Section 3.1 of the main report

[13] See http://www.ramesesproject.org/

No.	Standard	Relevant section of report or annexes
product evaluated		
10. Describe and justify the evaluation design	A description and justification of the evaluation design (i.e. the account of what was planned, done and why) should be included, at least in summary form or as an appendix, in the document which presents the main findings. If this is not done, the omission should be justified and a reference or link to the evaluation design given. It may also be useful to publish or make freely available (e.g. online on a website) any original evaluation design document or protocol, where they exist	See Annex 3
11. Data collection methods	Describe and justify the data collection methods – which ones were used, why and how they fed into developing, supporting, refuting or refining programme theory. Provide details of the steps taken to enhance the trustworthiness of data collection and documentation	See Annex 3.4
12. Recruitment process and sampling strategy	Describe how respondents to the evaluation were recruited or engaged and how the sample contributed to the development, support, refutation or refinement of programme theory	See Annex 3.4 and Annex 8
13. Data analysis	Describe in detail how data were analysed. This section should include information on constructs that were identified, process of analysis, how the programme theory was further developed, supported, refuted and refined, and (where relevant) how analysis changed as the evaluation unfolded.	See Annex 3.4, 3.7 and 4.
Results		
14. Details of participants	Report (if applicable) who took part in the evaluation, the details of the data they provided and how the data was used to develop, support, refute or refine programme theory	See Annex 3.4
15. Main findings	Present the key findings, linking them to contexts, mechanisms and outcome configurations. Show how they were used to further develop, test or refine the programme theory.	See Sections 5-7 of the report, with further detail in Annex 4
Discussion		
16. Summary of findings	Summarise the main findings with attention to the evaluation questions, purpose of the evaluation, programme theory and intended audience	Summaries of the main findings are included throughout the report in tables and summary boxes, and in the overall conclusions
17. Strengths, limitations and future directions	Discuss both the strengths of the evaluation and its limitations. These should include (but need not be limited to): (1) consideration of all the steps in the evaluation processes and (2) comment on the adequacy, trustworthiness and value of the explanatory insights which emerge. In many evaluations, there will be an expectation to provide guidance on future directions for the programme, policy or initiative, its implementation and/or design. The particular implications arising from the realist nature of the findings should be reflected in these discussions	See Section 2.4 of the main report
18. Comparison with existing literature	Where appropriate, compare and contrast the evaluation's findings with the existing literature on similar programmes policies or initiatives	This is done throughout the report in 'insights from the literature' boxes
19. Conclusion and recommendations	List the main conclusions that are justified by the analyses of the data. If appropriate, offer recommendations consistent with a realist approach	See conclusions

No.	Standard	Relevant section of report or annexes
20. Funding and conflict of interest	State the funding source (if any) for the evaluation, the role played by the funder (if any) and any conflicts of interests of the evaluators.	See Section 1 of the main report. Further details on the evaluation team are contained in Annex 3.10

12. References

Befani, B. and Mayne, J. (2014). Process Tracing and Contribution Analysis: A Combined Approach to Generative Causal Inference for Impact Evaluation. *IDS Bulletin, 45(*6): 17–36. Link.

DFID. (2009). Political Economy Analysis: How to Note. A DFID Practice Paper. Link.

DFID. (2010). How To Note on Capacity Building in Research. Practice Paper. London: DFID.

du Toit, A. (2012). Making Sense of 'Evidence': Notes on the Discursive Politics of Research and Pro-Poor Policymaking. Working Paper 21. Cape Town: Institute for Poverty, Land and Agrarian Studies. Link.

Funnell, S.C. and Rogers, P.J. (2011). Purposeful Program Theory: Effective Use of Theories of Change and Logic Models. San Francisco, CA: Jossey-Bass

Hudson, D., Marquette, H. and Waldock, S. (2016). Everyday Political Analysis, Developmental Leadership Programme, University of Birmingham. Link.

IDRC. (2014). Lasting Change – Strengthening Capacity to Improve Health Systems. Ottawa: IDRC. Link.

Lemire, S.T., Nielsen, S.B. and Dybdal, L. (2012). Making contribution analysis work: A practical framework for handling influencing factors and alternative explanations. *Evaluation, 18*(3).

Mayne, J. (2008). Contribution analysis: An approach to exploring cause and effect. International Learning and Change (ILAC) Brief, 16.

Mayne, J. (2011). Addressing cause and effect in simple and complex settings through contribution analysis. In R. Schwartz, K. Forss and M. Marra (eds), *Evaluating the Complex*. Piscataway, NJ: Transaction.

Mayne, J. (2012). Contribution analysis: Coming of age? Evaluation, 18(3): 270–80.

Michaelis, C., & Westhorp, G. (2016). International Climate Fund Evaluation Evidence Framework, Final Draft.

Moncrieffe, J. and Luttrell, C. (2005). An Analytical Framework for Understanding the Political Economy of Sectors and Policy Arenas. Overseas Development Institute, 1–36.

Noblit, G. and Hare, R. (1988). *Meta-Ethnography: Synthesizing Qualitative Studies*. London: Sage.

Nutley, S. M., Davies, H. and Walter, I. (2002). Evidence Based Policy and Practice: Cross Sector Lessons From the UK. Working Paper 9. ESRC UK Centre for Evidence Based Policy and Practice.

Pawson, R. (2006). Evidence-based Policy: A Realist Perspective. London: SAGE Publications.

Pawson, R. and Tilley, N. (1997). *Realistic Evaluation*. London: Sage.

Pawson, R. and Tilley, N. (2004). *Realist Evaluation*. Link.

Poole, A. (2011) 'Political Economy Assessments at Sector and Project Levels', How-To Note. Washington DC: World Bank. Link.

Pope, C., Mays, N. and Popay, J. (2007). Synthesising Qualitative And Quantitative Health Evidence: A Guide To Methods. McGraw-Hill International.

Punton, M., Brown, C., Echt, L., Fisher, C., Lloyd, R., Morton, S. and Vogel, I. (2016a). How Can Capacity Development Promote Evidence-Informed Policymaking? Literature Review for the BCURE programme. Brighton. Link.

Punton, M., Vogel, I. and Lloyd, R. (2016b). Reflections from a realist evaluation in progress: scaling ladders and stitching theory. *CDI Practice Paper*, (18). Link.

Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., & Befani, B. (2012). Broadening the Range of Designs and Methods for Impact Evaluations. DFID Working Paper, 38(April 2012).

Sutcliffe, S. and Court, J. (2005). *Evidence-Based Policymaking: What Is It? How Does It Work? What Relevance for Developing Countries?* London: ODI.

Waldman, T. (2014). The Use of Statebuilding Research in Fragile Contexts: Evidence from British Policymaking in Afghanistan, Nepal and Sierra Leone. Journal of Intervention and Statebuilding, 8(2–3): 149–72. Link.

Westhorp, G. (2014). Realist Impact Evaluation: An Introduction. Methods Lab. London: ODI. Link.

Wong, G., Westhorp, G., Manzano, A., Greenhalgh, J., Jagosh, J. and Greenhalgh, T. (2016). RAMESES II reporting standards for realist evaluations. *BMC Medicine, 14*(1), 96. Link.

Wimbush, E., Montague, S. and Mulherin, T. (2012). Applications of contribution analysis to outcome planning and impact evaluation. Evaluation, 18(3), 310–29. Link.

Wong, G., Westhorp, G., Pawson, R. and Greenhalgh, T. (2013). *Realist Synthesis: Rameses Training Materials* (misc). London: Health Research Health Services and Delivery Research Program. Link.

Wong, G., Westhorp, G., Manzano, A., Greenhalgh, J., Jagosh, J. and Greenhalgh, T. (2016). RAMESES II Reporting Standards for Realist Evaluations. *BMC Medicine, 14*(1), 96. Link.

Yin, R. K. (2003). *Case Study Research: Design and Methods*. California: Sage.